Quantitative study of piecewise deterministic Markov processes for modeling purposes

Florian Bouguet

ANNÉE 2016

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Bretagne Loire

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Mathématiques et applications

École doctorale Matisse

présentée par

Florian Bouguet

Préparée à l'IRMAR – UMR CNRS 6625
Institut de recherche mathématique de Rennes
U.F.R. de Mathématiques

Étude quantitative de processus de Markov déterministes par morceaux issus de la modélisation

Thèse soutenue à Rennes
le 29 juin 2016
devant le jury composé de :

Bernard BERCU
Professeur à l'Université Bordeaux 1 / *Examinateur*

Patrice BERTAIL
Professeur à l'Université Paris Ouest Nanterre La Défense / *Examinateur*

Jean-Christophe BRETON
Professeur à l'Université Rennes 1 / *Co-directeur de thèse*

Patrick CATTIAUX
Professeur à l'Université Toulouse 3 / *Examinateur*

Anne GÉGOUT-PETIT
Professeur à l'Université de Lorraine / *Rapporteur*

Hélène GUÉRIN
Maître de conférence à l'Université Rennes 1 / *Examinatrice*

Eva LÖCHERBACH
Professeur à l'Université Cergy-Pontoise / *Rapporteur*

Florent MALRIEU
Professeur à l'Université de Tours / *Directeur de thèse*

Résumé

L'objet de cette thèse est d'étudier une certaine classe de processus de Markov, dits déterministes par morceaux, ayant de très nombreuses applications en modélisation. Plus précisément, nous nous intéresserons à leur comportement en temps long et à leur vitesse de convergence à l'équilibre lorsqu'ils admettent une mesure de probabilité stationnaire. L'un des axes principaux de ce manuscrit de thèse est l'obtention de bornes quantitatives fines sur cette vitesse, obtenues principalement à l'aide de méthodes de couplage. Le lien sera régulièrement fait avec d'autres domaines des mathématiques dans lesquels l'étude de ces processus est utile, comme les équations aux dérivées partielles. Le dernier chapitre de cette thèse est consacré à l'introduction d'une approche unifiée fournissant des théorèmes limites fonctionnels pour étudier le comportement en temps long de chaînes de Markov inhomogènes, à l'aide de la notion de pseudo-trajectoire asymptotique.

Mots-clés : Processus de Markov déterministes par morceaux ; Ergodicité ; Méthodes de couplage ; Vitesse de convergence ; Modèles de biologie ; Théorèmes limites fonctionnels

Abstract

The purpose of this Ph.D. thesis is the study of piecewise deterministic Markov processes, which are often used for modeling many natural phenomena. Precisely, we shall focus on their long time behavior as well as their speed of convergence to equilibrium, whenever they possess a stationary probability measure. Providing sharp quantitative bounds for this speed of convergence is one of the main orientations of this manuscript, which will usually be done through coupling methods. We shall emphasize the link between Markov processes and mathematical fields of research where they may be of interest, such as partial differential equations. The last chapter of this thesis is devoted to the introduction of a unified approach to study the long time behavior of inhomogeneous Markov chains, which can provide functional limit theorems with the help of asymptotic pseudotrajectories.

Keywords: Piecewise deterministic Markov processes; Ergodicity; Coupling methods; Speeds of convergence; Biological models; Functional limit theorems

REMERCIEMENTS

Tout d'abord, je tiens à exprimer ma profonde gratitude à mes directeurs de thèse, Florent Malrieu et Jean-Christophe Breton, pour m'avoir donné envie de faire de la recherche, pour leur disponibilité, pour leurs nombreux conseils et encouragements, en bref pour m'avoir guidé pendant ces trois années. Un grand merci également à Anne Gégout-Petit et Eva Löcherbach, pour avoir accepté de rapporter ma thèse, pour leur relecture attentive et pour l'intérêt qu'elles ont porté à mon travail. Enfin je remercie Bernard Bercu, Patrice Bertail, Patrick Cattiaux et Hélène Guérin de me faire l'honneur de leur présence dans mon jury, et plus particulièrement Patrice et Hélène pour leur soutien tout au long de ces trois années.

D'autre part, je tiens à remercier très chaleureusement Bertrand Cloez pour les nombreuses et fructueuses discussions que nous avons eues ensembles depuis la Suisse, pour ses invitations à Montpellier et pour m'avoir fait profiter de son expérience de la recherche. Egalement, un grand merci à Michel Benaïm pour m'avoir accueilli à Neuchâtel, ainsi que pour son enthousiasme et sa curiosité mathématique. J'en profite également pour remercier Julien Reygner, Fabien Panloup et Christophe Poquet, certes pour avoir écrit avec moi un acte de conférence, mais aussi pour les diverses discussions que nous avons pu avoir au cours de ces années. D'autre part, je remercie chaleureusement Romain Azaïs, Anne Gégout-Petit et Aurélie Muller-Gueudin pour me donner l'occasion de travailler avec eux l'année prochaine à Nancy, ainsi que Fabien Panloup, Bertrand Cloez, Guillaume Martin, Tony Lelièvre, Pierre-André Zitt et Mathias Rousset pour avoir accepté de constituer des dossiers de post-doc avec moi. Pour finir, un remerciement général aux membres de l'ANR Piece ainsi qu'à toutes les personnes m'ayant invité à exposer mes travaux, et tous les doctorants et jeunes chercheurs qui sont venus parler au séminaire Gaussbusters.

Bien évidemment, je souhaite aussi remercier les rennais en général, et les chercheurs de l'IRMAR en particulier : je pense notamment à Mihai, Hélène, Nathalie, Jürgen, Ying, Guillaume, Rémi, Stéphane et Dimitri. Ma reconnaissance va également à tous les (autres) professeurs de mathématiques que j'ai pu avoir au cours de ma scolarité, tant à l'ENS qu'auparavant, pour m'avoir donné le goût de la logique et des théorèmes, ainsi que les (autres) enseignants dont j'ai eu l'occasion de gérer les TD ou TP durant ces trois années. De même, je dois un grand merci à toute l'équipe administrative de l'université pour son travail, sans lequel je n'aurais pas fait grand chose au cours de ma thèse : merci en particulier à Emmanuelle, Chantal, Marie-Aude, Hélène (encore), Marie-Annick, Xhensila et Olivier. Je tiens également à remercier tous mes cobureaux successifs (par ordre de durée, en tout cas sur le papier) : Margot, Camille, Felipe, Tristan et Damien, ainsi qu'à titre exceptionnel Blandine et Richard. Il ne me reste plus qu'à remercier tous les doctorants (actuels ou exilés) de l'IRMAR pour leur bonne humeur et pour les longs moments partagés au RU et dans le bureau 202/1. Notamment, merci à Jean-Phi pour nos marathons, à Ophélie pour ses chocolats, à Julie pour ses recettes (et son humour non-assumé), à Hélène (et encore) pour son humour, à Hélène (et toujours) pour ses talons avant-coureurs, à Mac pour nos chambres au CIRM, à Blandine pour ses gâteaux post-séminaire, à Tristan pour servir si souvent le café[1] (et pour avoir agi en *loucedé*), à Richard pour son mariage, à Maxime pour ce jeu frustrant dont j'ai oublié le nom, à Vincent pour ses bougies d'anniversaire, à Yvan pour ses goûts sûrs, à Axel pour donner le coup d'envoi du pot tout à l'heure, à Néstor pour ses cours d'espagnol, à Arnaud pour ses mails toujours utiles, à Benoit pour chanter du Disney, à Alexandre pour son flamant rose[2], à Christian pour son hébergement chaleureux, à Andrew pour ses ingrédients bizarres, à Adrien pour tous ces matchs de tennis qu'on aurait dû faire, à Marine pour rentrer chez elle en courant, à Cyril pour sa comédie musicale, à Olivier pour sa prudence au tarot, à Renan pour nous narguer sur les réseaux sociaux, à Pierre-Yves pour le Perudo, à Marie pour m'avoir refilé le séminaire, à Salomé pour la bouteille, à Tristan pour les tasses, et à Damien et Charles pour leurs (longues) conversations. Merci de m'avoir accompagné et supporté (dans tous les sens du terme) dans cette aventure.

Ayant évoqué Neuchâtel plus haut, j'aimerais remercier Mireille, Carl-Erik, Edouard et tous les mathématiciens neuchâtelois pour leur splendide accueil dans leur belle (quoique sous le brouillard en automne) ville. Dans le même ordre d'idée, merci aux chercheurs tourangeaux pour m'avoir si bien reçu lors de mes visites à l'université François Rabelais. Au cours de ma thèse, j'ai eu l'occasion de rencontrer de nombreux doctorants de tous horizons, que j'aimerais remercier pour leur sympathie et nos nombreuses discussions plus ou moins sérieuses. Citons mes deux biloutes Geoffrey et Benjamin, Pierre (Monmarché/Houdebert/Hodara), Olga, Thibaut, Ludo, Marie, Claire, Eric, Gabriel, Aline, Alizée, et même Marie-Noémie.

Pour sortir du cadre des maths, je dois beaucoup à mes amis (bretons et autres) qui me soutiennent depuis de très nombreuses années. Commençons par remercier Paul, Cricri, Basoune, Coco et Andéol (entre 5 et 7 ans, notre amitié rentre en primaire), puis viennent Max, Flow, Mika, Playskool, Yoyo, Dave, Elodie, Seb, Sarah, Jean, Amélie, Florian, Gaël, Solenn (entre 4 et 12 ans, le cap du collège) et Antoine, Florent et Vivien (entre 14 et 18 ans, la crise d'adolescence). Ca ne nous rajeunit pas, ma bonne dame !

Pour finir, un grand merci à toute ma famille pour leur soutien, et une pensée à ceux qui ont disparu trop tôt. Ayant gardé les meilleurs pour la fin, merci à vous Anne-Marie et Jean-Luc pour vos encouragements et votre affection constants, et tout ce que vous avez fait pour moi. Et merci à toi Elodie, pour ton aide, ton réconfort et ta présence.

[1]Dans la tasse. En général.

[2]Quand on voit sa tête, c'est pas étonnant…

TABLE DES MATIÈRES

- Sir, the possibility of successfully navigating
an asteroid field is approximately 3,720 to 1.
- Never tell me the odds.

CHAPITRE 0

AVANT-PROPOS

Dans cette thèse de doctorat, nous nous intéresserons aux dynamiques d'un certain type de processus stochastiques, les processus de Markov déterministes par morceaux, ou *Piecewise Deterministic Markov Process* (PDMP). Les PDMP ont été historiquement introduits par Davis dans [Dav93] et ont depuis été intensivement étudiés, car ils apparaissent naturellement dans de nombreux domaines des sciences ; citons par exemple l'informatique, la biologie, la finance, l'écologie, etc.

Un PDMP est un processus suivant une évolution déterministe (typiquement, régie par une équation différentielle), mais qui change de dynamique à des instants aléatoires. Ces sauts, comme on les appelle, peuvent survenir à des instants aléatoires, et leurs mécaniques (déclenchement et direction de saut) peuvent dépendre de l'état actuel du processus. Un outil-clé dans l'étude des PDMP est leur générateur infinitésimal ; il est facile de lire la dynamique d'un processus sur son générateur, où sont transcrits à la fois son comportement inter-sauts, et toute la mécanique du saut. De manière grossière, on pourrait séparer les PDMP en deux catégories. D'un côté, on rencontre des processus possédant uniquement une composante spatiale, qui auront des trajectoires discontinues. C'est cette composante spatiale qui saute, et on observe alors ce saut directement sur la trajectoire du processus. Ces processus modélisent de très nombreux phénomènes, et nous suivrons l'exemple d'un modèle intervenant en pharmacocinétique (étude de l'évolution d'une substance chimique après administration dans l'organisme). D'un autre côté, de nombreux PDMP sont décrits à l'aide de composantes spatiales et d'une composante discrète, cette dernière servant à caractériser le flot (et donc la dynamique) suivi par le processus. Il est alors courant d'obtenir des trajectoires continues, mais changeant brutalement lorsque le flot lui-même change. Ces processus permettent souvent de modéliser des phénomènes déterministes en milieu aléatoire. Si deux échelles temporelles se distinguent nettement dans ces phases, on retrouvera éventuellement des PDMP de la première catégorie en assimilant les phases rapides à des sauts.

Un problème récurrent dans l'étude de processus stochastiques est leur comportement asymptotique. En effet, il est fréquent de se retrouver en situation d'ergodicité, la loi du processus convergeant alors vers une loi de probabilité dite stationnaire. De nombreux problèmes se soulèvent alors d'eux-mêmes : déterminer la vitesse de convergence à l'équilibre, qui dépend bien souvent de la métrique choisie, déterminer, simuler ou simplement obtenir des informations sur la loi stationnaire, etc. Le monde des processus de Markov déterministes par morceaux est riche et vaste, et la littérature abonde en ce qui concerne leur vitesse de convergence à l'équilibre. Dans ce manuscrit, nous traiterons particulièrement de manière poussée le critère de Foster-Lyapunov et de nombreuses méthodes de couplage. Il est globalement difficile d'obtenir des vitesses de convergence explicites et satisfaisantes dans un cadre général, et c'est pourquoi nous ferons apparaître au maximum les liens entre les différents PDMP apparaissant dans la modélisation de phénomènes physiques.

Ce manuscrit est découpé en quatre chapitres. Dans une première partie, nous replacerons la thèse dans son contexte et décrirons les problématiques mises en jeu. Nous rappellerons les notions de base nécessaires à la bonne compréhension du reste de ce mémoire. Dans une seconde partie, nous étudierons la vitesse de mélange d'une classe de processus de Markov déterministes par morceaux particulièrement utilisés dans des modèles de pharmacocinétique, dont les instants de sauts sont régis par un processus de renouvellement. Le troisième chapitre regroupe des résultats plus isolés sur les PDMP. Il y sera notamment question de processus *shot-noise*, d'équations de croissance/fragmentation et de retournement du temps. Enfin, le dernier chapitre présente une méthode unifiée pour approcher une chaîne de Markov inhomogène à l'aide d'un processus de Markov homogène, et pour déduire des propriétés asymptotiques de la première à partir de celles du second. Dans tout ce manuscrit, un exemple simple de processus de Markov fera office de fil conducteur pour comprendre les phénomènes-clés mis en évidence.

Les simulations ont été générées avec Scilab, et les illustrations avec TikZ. Ce mémoire de thèse a quant à lui été principalement généré à partir des articles suivants :

- Florian Bouguet. Quantitative speeds of convergence for exposure to food contaminants. *ESAIM Probab. Stat.*, 19 :482-501, 2015.
- Florian Bouguet, Florent Malrieu, Fabien Panloup, Christophe Poquet, and Julien Reygner. Long time behavior of Markov processes and beyond. *ESAIM : Proc. Surv.*, 51 :193-211, 2015.
- Michel Benaïm, Florian Bouguet, and Bertrand Cloez. Ergodicity of inhomogeneous Markov chains through asymptotic pseudotrajectories. *ArXiv e-prints*, January 2016.

CHAPITRE 1

INTRODUCTION GÉNÉRALE

Dans ce chapitre, nous posons les bases nécessaires pour comprendre l'ensemble de ce manuscrit. Nous reviendrons notamment en détail sur les notions de processus de Markov déterministe par morceaux, de générateur infinitésimal, d'ergodicité et de couplage, ainsi que de nombreuses notions voisines utiles pour comprendre le tout. On fera régulièrement référence à un exemple-jouet au comportement simple, introduit à la Remarque 1.1.1 et issu de problèmes de risque alimentaire, pour illustrer des notions importantes tout au long du chapitre.

Commençons par introduire quelques notations :

- $\mathscr{M}_1(\mathbb{X})$ est l'ensemble des mesures de probabilité sur un espace $\mathbb{X}$.
- $\mathscr{L}(X)$ est la distribution de probabilité d'un objet aléatoire X (typiquement un vecteur aléatoire ou un processus stochastique), et $\mathrm{Supp}(\mathscr{L}(X))$ son support. On écrira aussi $X \sim \mathscr{L}(X)$.
- δ_x est la mesure de Dirac en $x \in \mathbb{R}^d$.
- $\mathscr{C}_b^N(\mathbb{R}^d)$ est l'ensemble des fonctions de $\mathscr{C}^N(\mathbb{R}^d)$ (N fois continûment différentiables) telles que $\sum_{j=0}^N \|f^{(j)}\|_\infty < +\infty$, pour $N \in \mathbb{N}$.
- $\mathscr{C}_c^N(\mathbb{R}^d)$ est l'ensemble des fonctions $\mathscr{C}^N(\mathbb{R}^d)$ à support compact, pour $N \in \mathbb{N}$ ou $N = +\infty$.
- $\mathscr{C}_0^0(\mathbb{R}^d) = \{f \in \mathscr{C}^0(\mathbb{R}^d) : \lim_{\|x\|\to\infty} f(x) = 0\}$.
- $x \wedge y = \min(x, y)$ et $x \vee y = \max(x, y)$ pour tous $x, y \in \mathbb{R}$.

Lorsqu'il n'y aura pas d'ambiguïté, l'espace sur lequel on considère les mesures de probabilité ou les fonctions ne sera pas toujours indiqué.

1.1 Processus de Markov

1.1.1 Semi-groupe et générateur

Intéressons-nous maintenant aux processus de Markov, qui représentent le cœur de cette thèse. Le lecteur intéressé par de plus amples détails pourra consulter par exemple [EK86] ou [Kal02]. On commence par se donner un processus de Markov homogène en temps $(X_t)_{t\geq 0}$, à valeurs dans $\mathbb{R}^d$[1], et à trajectoire continue à droite, limite à gauche (càdlàg) presque sûrement (p.s.) On peut définir son semi-groupe $(P_t)_{t\geq 0}$ comme la famille d'opérateurs tels que

$$P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x],$$

pour n'importe quelle fonction f mesurable bornée. Dans la suite, on travaillera sur l'espace $\mathscr{C}_0^0$, ce qui sera justifié dans quelques lignes. Il est à noter que

$$\|P_t f\|_\infty \leq \|f\|_\infty.$$

Dans ce manuscrit, nous considérerons des semi-groupes dits de Feller, c'est-à-dire que pour toute fonction $f \in \mathscr{C}_0^0$, $P_t f \in \mathscr{C}_0^0$ et $\lim_{t\to 0} \|P_t f - f\|_\infty = 0$. Il est à noter que si son semi-groupe bénéficie de la propriété de Feller, le processus X vérifie la propriété de Markov forte. Si $\mu \in \mathscr{M}_1$, on écrira volontiers

$$\mu(f) = \int_{\mathbb{R}^d} f(x)\mu(dx), \quad \mu P_t = \mathscr{L}(X_t|X_0 \sim \mu).$$

Il est facile de vérifier que

$$\mu(P_t f) = \mu P_t(f), \quad P_{t+s} = P_t P_s,$$

cette dernière égalité étant appelée relation de Chapman-Kolmogorov (justifiant l'appellation *semi-groupe*). Cette relation peut aussi être vue comme un semi-flot sur l'espace $\mathscr{M}_1$ des lois de probabilités, comme ce sera le cas au Chapitre 4.

Un outil fondamental dans l'étude des processus de Markov est le générateur infinitésimal. Rigoureusement, on le définit comme étant l'opérateur agissant sur les fonctions f telles que $\lim_{t\to 0} \|t^{-1}(P_t f - f) - \mathcal{L} f\|_\infty = 0$. On note $\mathcal{D}(\mathcal{L})$ son domaine, autrement dit l'ensemble des fonctions pour lesquelles cette limite est vérifiée ; ce domaine est dense dans $\mathscr{C}_0^0$. Alors, si $f \in \mathcal{D}(\mathcal{L})$, $P_t f \in \mathcal{D}(\mathcal{L})$ et vérifie

$$\partial_t P_t f = \mathcal{L} P_t f = P_t \mathcal{L} f, \quad P_t f = f + \int_0^t \mathcal{L} P_s f ds.$$

Il est à noter qu'un semi-groupe, et donc la dynamique d'un processus de Markov, est entièrement caractérisé par la donnée de ce générateur et de son domaine. De plus, il est généralement explicite et facilite les calculs, à l'inverse du semi-groupe qui n'est souvent pas accessible directement. Tout au long de ce manuscrit, c'est souvent le générateur qui sera donné afin de définir la dynamique d'un processus de Markov.

[1] Plus généralement, on pourrait travailler dans un espace polonais muni de sa tribu borélienne.

Remarque 1.1.1 (*Un exemple introductif : le processus pharmacocinétique*[2]) : Soient $(\Delta T_n)_{n\geq 1}$ et $(U_n)_{n\geq 1}$ des suites de variables aléatoires indépendantes et identiquement distribuées, mutuellement indépendantes, de lois exponentielles respectives $\mathscr{E}(\lambda)$ et $\mathscr{E}(\alpha)$. Considérons la chaîne de Markov $(\widetilde{X}_n)$ à valeurs dans $\mathbb{R}_+$ telle que, pour $n \in \mathbb{N}^*$,

$$\widetilde{X}_{n+1} = \widetilde{X}_n \exp\left(-\theta \Delta T_{n+1}\right) + U_{n+1}.$$

Notons $T_n = \sum_{k=1}^n \Delta T_k$ et $(X_t)_{t\geq 0}$ le processus stochastique tel que

$$X_t = \sum_{n=0}^{\infty} \widetilde{X}_n \exp(-\theta(t-T_n)) \mathbb{1}_{T_n \leq t < T_{n+1}}.$$

Typiquement, X_t décroit exponentiellement suivant l'équation différentielle $\partial_t y = -\theta y$ et effectue des sauts additifs de hauteur U_n aux instants T_n (voir Figure 1.1.1). Le processus (X_t) est alors un processus de Markov, dit déterministe par morceaux, à trajectoires càdlàg et de générateur infinitésimal

$$\mathcal{L}f(x) = -\theta x f'(x) + \lambda \int_0^{\infty} [f(x+u) - f(x)] \alpha \mathrm{e}^{-\alpha u} du. \tag{1.1.1}$$

Il sera souvent fait appel à cet exemple simple, voire simpliste, pour illustrer les propos de ce manuscrit. On peut l'imaginer, et ce sera le contexte du Chapitre 2, comme modélisant la quantité de contaminant alimentaire dans le corps à l'instant t, voir par exemple [BCT08]. Les dates T_n représentent alors les instants d'ingestion d'une quantité U_n de nourriture, entre lesquels le corps du sujet essaie d'éliminer les substances chimiques indésirables. Pour être au plus proche de la réalité, il sera intéressant de modifier les lois des variables aléatoires régissant le temps d'inter-ingestion, la quantité ingérée ou l'élimination métabolique.

Comme nous le verrons en Section 1.1.2, il est possible de lire la dynamique d'un processus de Markov à travers son générateur. En attendant, démontrons rapidement pourquoi c'est bien ce générateur que l'on obtient. Pour f suffisamment régulière (disons dans $\mathscr{C}_b^2$), on a, conditionnellement à $\{X_0 = x\}$,

$$\begin{aligned}
\mathbb{E}\left[f(X_t)\mathbb{1}_{t<T_1}\right] &= f(x\mathrm{e}^{-\theta t})\mathrm{e}^{-\lambda t} = f(x) - t\left(\theta x f'(x) + \lambda f(x)\right) + o(t), \\
\mathbb{E}\left[f(X_t)\mathbb{1}_{T_1\leq t<T_2}\right] &= \int_0^t \mathbb{E}\left[f(X_t) | s = T_1 \leq t < T_2\right] \mathbb{P}\left(T_1 \leq t < T_2 | T_1 = s\right) \lambda \mathrm{e}^{-\lambda s} ds \\
&= \int_0^t \mathbb{E}\left[f\left((x\mathrm{e}^{-\theta s} + U_1)\mathrm{e}^{-\theta(t-s)}\right)\right] \mathbb{P}(\Delta T_2 > t-s) \lambda \mathrm{e}^{-\lambda s} ds \\
&= \int_0^t \mathbb{E}\left[f\left((x\mathrm{e}^{-\theta s} + U_1)\mathrm{e}^{-\theta(t-s)}\right)\right] \lambda \mathrm{e}^{-\lambda t} ds \\
&= \lambda t \mathrm{e}^{-\lambda t} \mathbb{E}\left[f(x+U_1)\right] + o(t) = \lambda t \mathbb{E}\left[f(x+U_1)\right] + o(t), \\
\mathbb{E}\left[f(X_t)\mathbb{1}_{T_2\leq t}\right] &= \iint_{s_1 \leq s_2 \leq t} \mathbb{E}\left[f(X_t)\mathbb{1}_{T_2 \leq t} | s_1 = T_1, s_2 = T_2 \leq t\right] \lambda^2 \mathrm{e}^{-\lambda(s_2-s_1)} \mathrm{e}^{-\lambda s_1} ds_1 ds_2 \\
&= o(t).
\end{aligned}$$

[2]La pharmacocinétique désigne l'étude de la dynamique de substances chimiques dans le corps.

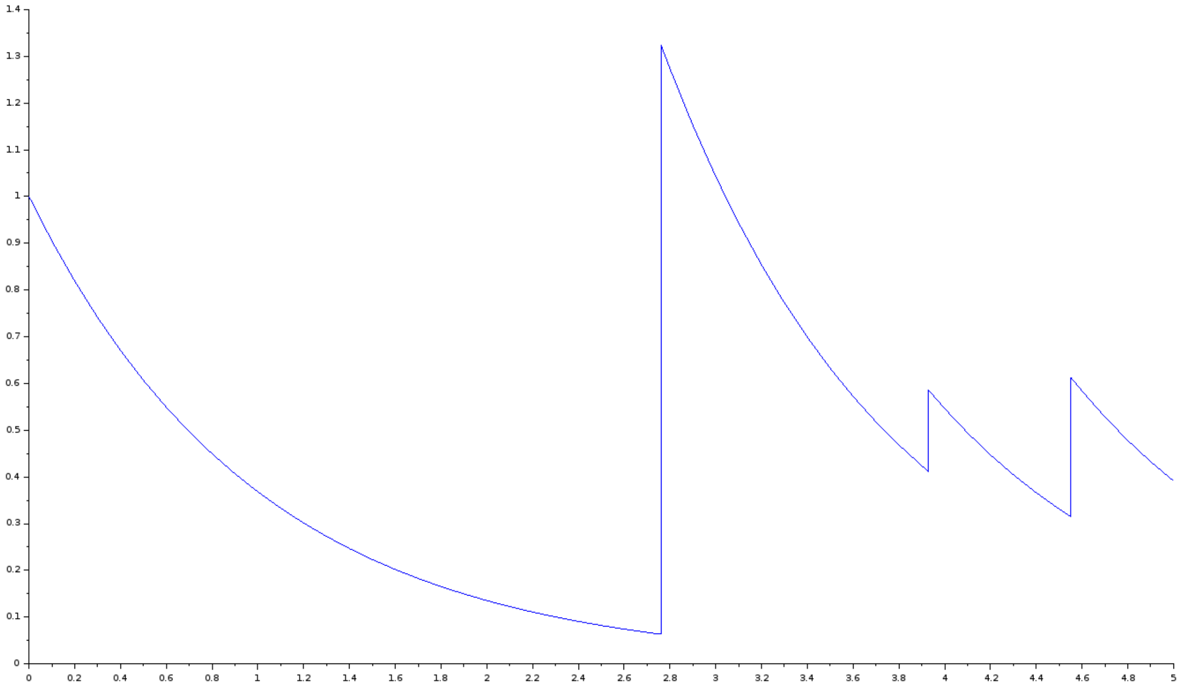


FIGURE 1.1.1 – Simulation du processus généré par (1.1.1), pour $\theta = 1, \lambda = 0.5, \alpha = 2$.

On a donc

$$\begin{aligned}\mathcal{L}f(x) &= \lim_{t \to 0^+} \frac{P_t f(x) - f(x)}{t} = -\theta x f'(x) + \lambda \mathbb{E}[f(x + U_1)] - \lambda f(x) \\ &= -\theta x f'(x) + \lambda \int_0^\infty [f(x+u) - f(x)] \alpha \mathrm{e}^{-\alpha u} du.\end{aligned}$$

$\diamondsuit$

Terminons cette section en évoquant la notion de mesure stationnaire. Une loi $\pi \in \mathscr{M}_1$ est dite stationnaire, ou invariante, si, pour tout $t \geq 0$, $\pi P_t = \pi$ ou, de manière équivalente, $\pi(\mathcal{L}f) = 0$ pour toute fonction $f \in L^2(\pi)$. Cela signifie que si le processus X démarre sous la loi π (i.e. $X_0 \sim \pi$), alors il gardera cette loi à tout temps. De nombreux processus de Markov possèdent une unique mesure invariante vers laquelle ils convergent en temps long, dans un sens à préciser ; c'est ce que l'on appelle l'ergodicité. Que l'on voit cette notion comme une manière de générer une variable aléatoire sous π, ce qui est le principe des méthodes *Markov Chain Monte Carlo* (MCMC) (voir par exemple [GRS96, ADF01]) ou comme un comportement limite d'un phénomène à comprendre, la question de l'existence et de l'unicité de la mesure invariante est cruciale lorsque l'on s'intéresse à des processus de Markov. Dans ce manuscrit, nous étudierons plus précisément la convergence évoquée plus haut, et notamment la vitesse à laquelle elle s'opère, à l'aide de méthodes présentées en Section 1.2.

1.1.2 Processus de Markov déterministes par morceaux

Avant de parler de *Piecewise Deterministic Markov Process* (PDMP), nous allons d'abord introduire le taux de saut ; voir par exemple [Bon95]. On se donne donc une variable aléatoire ΔT positive presque sûrement, de fonction de répartition $F_{\Delta T}$ que l'on supposera à densité par rapport à la mesure de Lebesgue $f_{\Delta T}$. Le taux de saut de ΔT est la fonction λ valant 0 si $F_{\Delta T} = 1$ et sinon telle que, pour $t \geq 0$,

$$\lambda(t) = \frac{f_{\Delta T}(t)}{1 - F_{\Delta T}(t)} = \lim_{\varepsilon \to 0} \frac{\mathbb{P}(t \leq \Delta T \leq t + \varepsilon)}{\varepsilon \mathbb{P}(t \leq \Delta T)} = \partial_t \left(-\log(1 - F_{\Delta T}(t)) \right).$$

Intuitivement, il convient de penser à $\lambda(t)$ comme la volonté qu'a la variable aléatoire ΔT de se réaliser à l'instant t, sachant qu'elle ne s'est pas encore réalisée. Autrement dit, plus λ est élevé et plus la variable aléatoire aura tendance à être petite. On voit arriver le lien avec la loi exponentielle, qui est souligné par les relations suivantes :

$$F_{\Delta T}(t) = 1 - \exp\left(-\int_0^t \lambda(s)ds \right), \quad f_{\Delta T}(t) = \lambda(t) \exp\left(-\int_0^t \lambda(s)ds \right).$$

Il convient de remarquer que ΔT suit une loi exponentielle si, et seulement si, λ est constant ; dans ce cas, $\Delta T \sim \mathscr{E}(\lambda)$. C'est la fameuse propriété d'absence de mémoire de la loi exponentielle, et cette caractérisation fait que les inter-sauts exponentiels pour un PDMP sont un cadre confortable pour travailler, comme dans le cas du processus généré par (1.1.1) et comme on le verra ensuite au Chapitre 2. À noter que, si λ est majoré (resp. minoré par un réel strictement positif), il est alors possible de minorer (resp. majorer) ΔT stochastiquement par une loi exponentielle, ce qui sera très utile dans les méthodes de couplage qui suivent dans ce manuscrit. Enfin, remarquons que ΔT vérifie

$$\int_0^{\Delta T} \lambda(s)ds \sim \mathscr{E}(1),$$

ce qui est une relation classique pour simuler des réalisations de ΔT.

On peut maintenant s'intéresser aux processus de Markov déterministes par morceaux, introduits par [Dav93]. Trois éléments sont constitutifs d'un PDMP $(X_t)_{t\geq 0}$ évoluant dans $\mathbb{R}^d$: son champ de vecteurs $F : \mathbb{R}^d \to \mathbb{R}^d$ donnant le comportement déterministe entre les sauts, son taux de saut $\lambda : \mathbb{R}^d \to \mathbb{R}_+$ comme défini plus haut, et son noyau de saut $Q : \mathbb{R}^d \to \mathscr{M}_1$ définissant la façon dont le processus saute. Globalement, X évolue suivant le flot de F et saute avec des temps inter-sauts ΔT de taux λ, et suivant une loi $Q(x, dy)$ s'il saute de x à y. Comme annoncé plus haut, on peut lire toute la dynamique du PDMP dans son générateur infinitésimal

$$\mathcal{L}f(x) = \underbrace{F(x) \cdot \nabla f(x)}_{\text{comportement déterministe}} + \underbrace{\lambda(x)}_{\text{taux de saut}} \int_{\mathbb{R}^d} \underbrace{[f(y) - f(x)]Q(x, dy)}_{\text{direction de saut}}.$$

Nous ne démontrerons pas ce résultat ici, il s'obtient en suivant la méthode proposée à la Remarque 1.1.1 (voir aussi [Dav93, Théorème 26.14]). Dans la suite, nous supposerons que le nombre de sauts arrivant avant tout instant t est fini, ce qui nous assure de la non-explosion du processus (voir (24.8) dans [Dav93]).

Au contraire de nombreux processus diffusifs, les PDMP sont des processus de Markov non-réversibles et n'ont généralement pas d'effet régularisant :

- Si le champ de vecteur n'est pas nul, autrement dit si la dérive du processus entre ses sauts n'est pas constante, alors le PDMP ne sera pas réversible. Cela sera vu plus en détail au Chapitre 3.

- Si les temps d'inter-saut du processus ne sont pas bornés, et que $\mathscr{L}(X_0)$ est une mesure de Dirac, alors $\mathscr{L}(X_t)$ ne sera pas absolument continue par rapport à la mesure de Lebesgue. C'est ce qu'on appelle le manque d'effet régularisant, au contraire d'un processus diffusif qui satisferait une Equation Différentielle Stochastique (EDS) avec mouvement brownien, dont la loi au temps $t > 0$ chargera tout l'espace avec une mesure à densité.

Il est également à noter que de nombreux auteurs (par exemple [LP13b, ADGP14, ABG+14]) traitent le cas de processus évoluant dans des domaines où les PDMP sautent automatiquement s'ils touchent la frontière. Nous ne serons pas amenés à considérer de tels processus, car les modèles présentés dans ce manuscrit ne s'y prêtent pas, mais il est intéressant de noter que de nombreux résultats restent vrais dans ce cadre étendu. Notons enfin qu'on peut voir les PDMP comme des solutions d'EDS, sans mouvement brownien mais avec un processus de Poisson composé (voir par exemple [IW89, Fou02]). Si X un PDMP ayant pour générateur

$$\mathcal{L}f(x) = F(x) \cdot \nabla f(x) + \lambda(x) \int_{\mathbb{R}^d} [f(x + h(x,y)) - f(x)] Q(dy),$$

alors X est solution de l'EDS

$$X_t = X_0 + \int_0^t F(X_{s^-}) ds + \int_0^t \int_0^\infty \int_{\mathbb{R}^d} h(X_{s^-}, y) \mathbb{1}_{\{u \leq \lambda(X_{s^-})\}} N(ds\, du\, dy).$$

où N est une mesure de Poisson d'intensité $ds\, du\, Q(dy)$.

Les processus de renouvellement, que nous confondrons avec le *backward recurrence time process* défini dans [Asm03, Chapitre 5], sont un cas particulier de processus de Markov déterministes par morceaux. Il s'agit de processus évoluant dans $\mathbb{R}_+$, dont le générateur est de la forme

$$\mathcal{L}f(x) = f'(x) + \lambda(x)[f(0) - f(x)].$$

Ces processus croissent de manière linéaire, et tout leur aléa réside dans le taux de saut λ. Ils ont été très étudiés (citons [Lin92, Asm03, BCF15] pour les problématiques qui nous intéressent ici), et peuvent permettre de complexifier des modèles mathématiques pour les adapter un peu plus à la réalité. Ils autorisent la dynamique de saut d'un PDMP à dépendre du temps écoulé depuis le dernier saut, sans pour autant devoir étudier des processus de Markov inhomogènes en temps. Ces processus généralisent naturellement les processus de Poisson, ce qui sera l'une des motivations du Chapitre 2. En effet, dans le contexte de la pharmacocinétique, il n'est pas pertinent de supposer les temps d'ingestions comme étant distribués selon une loi exponentielle, mais plutôt avec un taux de défaillance croissant (comme le souligne [BCT10]).

La construction faite à la Remarque 1.1.1 à travers sa chaîne incluse (la suite de vecteurs aléatoires $(\widetilde{X}_n, T_n)_{n \geq 0}$) n'est pas anodine, et c'est même la façon classique de générer un PDMP. Dans le même ordre d'idées, il se trouve que l'on peut relier de nombreuses caractéristiques du processus (existence et unicité de la mesure invariante, stabilité, ergodicité...) à celles de certaines de ses chaînes incluses $(X_{\tau_n})_{n \geq 0}$ échantillonées de manière aléatoire. On pourra par exemple consulter [Cos90, CD08].

Remarque 1.1.2 (*Exemples de processus issus de la modélisation*) : Les processus de Markov déterministes par morceaux sont très largement utilisés en modélisation et en théorie du contrôle, et c'est ce que nous allons illustrer ici. Nous présentons ici quelques exemples directement issus de questions soulevées à la suite de la modélisation de phénomènes physiques, biologiques, etc. Cette liste n'est bien évidemment pas exhaustive et a été sélectionnée tant suivant mes goûts que suivant leur pertinence dans ce manuscrit. Le sujet a déjà été largement traité : par exemple, [RT15] liste plusieurs PDMP utilisés en biologie et [Mal15] présente plusieurs des modèles qui vont suivre. Citons également [All11], qui traite de très nombreux processus de Markov en temps discret ou continu.

i) Des questions de pharmacocinétique, comme nous l'avons vu à la Remarque 1.1.1, peuvent conduire à l'étude de processus évoluant sur $\mathbb{R}_+$ et ayant pour générateur

$$\mathcal{L}f(x) = -\theta x f'(x) + \lambda(x) \int_0^\infty [f(x+u) - f(x)] Q(du).$$

La généralisation et l'étude de ces processus est l'objet du Chapitre 2. Le lecteur intéressé par des questions de modélisation et de fondements biologiques pourra se référer à [GP82].

ii) Le processus *Transmission Control Protocol* (TCP) étudié par exemple dans [LvL08, CMP10, BCG$^+$13b], représente la quantité d'informations échangées sur un serveur. Cette quantité augmente linéairement jusqu'à ce qu'une saturation du système entraîne une division brutale par deux du flux de données ; cela revient à étudier un processus de Markov de générateur infinitésimal

$$\mathcal{L}f(x) = f'(x) + \lambda(x)[f(x/2) - f(x)].$$

Ce processus permet aussi de modéliser l'âge de bactéries, ou de cellules, et leur soudaine division en deux entités, comme dans [CDG12, DHKR15]. Le pendant analytique de ce phénomène et du modèle iii) est plus connu sous le nom d'équation de croissance/fragmentation. Nous aborderons ces processus en Section 3.2.1.

iii) Le capital d'une compagnie d'assurances, qui investit son argent et est de temps en temps amenée à fournir de grosses sommes d'argent à la suite de catastrophes naturelles, peut lui aussi être modélisé par un PDMP ; voir par exemple [KP11, AG15]. Alors, le générateur du processus est de la forme

$$\mathcal{L}f(x) = \theta x f'(x) + \lambda(x) \int_0^1 [f(xu) - f(x)] Q(x, du).$$

On verra au Chapitre 3 que ce processus peut-être vu comme le processus de pharmacocinétique cité plus haut retourné en temps, leurs dynamiques étant inversées. Bien évidemment, cela dépend fortement des caractéristiques des modèles, mais nous y reviendrons plus tard.

iv) Le processus processus on/off (ou processus de stockage), considéré par exemple dans [BKKP05], modélise par exemple la quantité d'eau dans un barrage qui suit deux régimes : ouvert et fermé. L'eau s'écoule ou s'emmagasine suivant le régime, ce qui conduit à étudier un processus $(X_t, I_t)_{t\geq 0}$ évoluant dans $(0,1) \times \{0,1\}$ de générateur infinitésimal

$$\mathcal{L}f(x,i) = (i-x)\theta\partial_x f(x,i) + \lambda[f(x,1-i) - f(x,i)].$$

Le processus (X_t) est attiré vers 0 et 1 alternativement, à vitesse exponentielle. Il s'agit du premier processus à flot changeant (ou *switching*) que nous rencontrons. Sa composante spatiale (X_t) est continue, et c'est la composante discrète (I_t), le régime en cours, qui indique à (X_t) le flot à suivre. La dynamique de ce processus est très simple, car le flot contracte exponentiellement, et il fera office d'exemple important pour introduire le "retournement en temps" de processus de Markov au Chapitre 3.

v) Le processus du télégraphe modélise l'évolution d'un micro-organisme sur la droite réelle, mouvement dont la vitesse varie suivant qu'il s'approche ou s'éloigne de l'origine (par exemple, s'il peut sentir la présence de nutriments en 0). On pourra consulter [FGM12, BR15b]. On obtient un processus de Markov $(X_t, I_t)_{t\geq 0}$ évoluant dans $\mathbb{R} \times \{-1,1\}$ dont la dynamique est dictée par le générateur

$$\mathcal{L}f(x,i) = if'(x) + [\alpha(x)\mathbb{1}_{\{xi\leq 0\}} + \beta(x)\mathbb{1}_{\{xi>0\}}][f(x,-i) - f(x,i)].$$

Si l'on suppose $\alpha \geq \beta$, la bactérie aura a priori plus envie de faire demi-tour si elle s'éloigne de l'origine.

vi) Il est intéressant d'introduire des flots changeants dans des modèles déterministes classiques, par exemple dans le cadre de la dynamique proie/prédateur modélisée par l'équation de Lotka-Volterra compétitive (voir par exemple [Per07]). Ces changements peuvent représenter l'évolution du climat, par exemple l'alternance des saisons. Comme dans les modèles iv) et v) cités plus haut, on considèrera un processus de Markov $(X_t, Y_t, I_t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \{0,1\}$ où (X_t, Y_t) suit alternativement (et de manière continue) les flots induits par deux équations de Lotka-Volterra compétitives, du type

$$\begin{cases} \partial_t X_t &= \alpha_{I_t} X_t(1 - a_{I_t} X_t - b_{I_t} Y_t) \\ \partial_t Y_t &= \beta_{I_t} Y_t(1 - c_{I_t} X_t - d_{I_t} Y_t) \end{cases},$$

et I_t est un processus à sauts sur un espace discret. Ici, X_t et Y_t représentent les populations de deux espèces en compétition. Ces PDMP sont notamment traités dans [BL14, MZ16]. Si $a_i < c_i$ et $b_i < d_i$, la saison i est favorable à l'espèce X. Suivant le rythme d'alternance des saisons, il se peut qu'une combinaison de saisons favorables à X lui soit finalement défavorable. On retrouve des phénomènes similaires avec les PDMP étudiés dans [BLBMZ14].

vii) L'expression des gènes, initiée par la transcription d'ARNm et suivie de sa traduction en protéines est couramment modélisée par des PDMPS : citons par exemple [YZLM14] et les références proposées à l'intérieur, et un modèle proche avec des flots changeants dans [BLPR07]. Si l'on note X et Y les concentrations respectives d'ARNm et de protéines, il a été observé que la transcription d'ARNm suit des pics d'activités (ou *bursting*) alors que la traduction en protéine s'opère de manière linéaire en la quantité d'ARNm. On obtient un processus $(X_t, Y_t)_{t\geq 0}$ suivant un générateur du type

$$\mathcal{L}f(x,y) = -\gamma_1 x\partial_x f(x,y) + (\lambda_2 x - \gamma_2 y)\partial_y f(x,y) \\ + \varphi(y)\int_0^\infty [f(x+u,y) - f(x,y)]H(du).$$

$\diamondsuit$

1.2 Comportement en temps long

Dans toute cette section, on cherche à donner un sens à la notion d'ergodicité mentionnée en Section 1.1.1. Dans quel sens la loi de X_t peut-elle converger vers une mesure stationnaire, et à quelle vitesse ?

1.2.1 Distances et couplages usuels

En probabilités, on dispose de nombreux types de convergence (presque sûre, en probabilité, dans L^p, etc.). La convergence qui nous intéresse ici est la plus faible d'entre toutes, la convergence en loi, de la loi d'un processus de Markov à l'instant t vers une mesure stationnaire, parfois appelée *équilibre*. On cherche donc à introduire des distances pour lesquelles la convergence implique la convergence en loi (ou convergence faible). Certaines d'entre elles sont particulièrement classiques, et le lecteur intéressé pourra consulter par exemple [Vil09]. Prenons $\mu, \nu \in \mathscr{M}_1$ et définissons la distance en variation totale[3] :

$$\|\mu - \nu\|_{TV} = \sup_{A\in\mathcal{B}(\mathbb{R}^d)} \{|\mu(A) - \nu(A)|\} = \frac{1}{2}\sup\{|\mu(\varphi) - \nu(\varphi)| : \|\varphi\|_\infty \leq 1\}. \quad (1.2.1)$$

Cette égalité est aisée à démontrer, et il est à noter que le supremum pourrait aussi être pris sur des fonctions seulement mesurables. On peut montrer que la distance en variation totale est issue d'une norme sur l'espace vectoriel des mesures signées, ce qui explique la notation $\|\cdot\|_{TV}$. Une autre distance elle aussi très utilisée est la distance de Wasserstein (pour laquelle on suppose que μ et ν admettent un moment d'ordre 1) :

$$W_1(\mu,\nu) = \sup\{|\mu(\varphi) - \nu(\varphi)| : \varphi \in \mathscr{C}^0, \varphi \text{ 1-lipschitzienne}\}. \quad (1.2.2)$$

Mais ces définitions paraissent bien analytiques, pour des distances sur un ensemble de lois de probabilités, comme étant des mesures agissant sur des fonctions. Après tout, les lois de probabilités ne sont-elles pas faites pour tirer des variables aléatoires ?

Nous introduisons donc une notion fondamentale pour la suite de ce manuscrit. On dit que $\gamma \in \mathscr{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$ est un couplage de μ et ν si, pour tout borélien A, $\gamma(A \times \mathbb{R}^d) = \mu(A)$ et $\gamma(\mathbb{R}^d \times A) = \nu(A)$. On demande donc à γ d'être une mesure

[3] Cette définition peut varier, à un facteur multiplicatif près.

sur l'espace produit, dont les marginales correspondent à μ et ν. Autrement dit, si l'on tire $(X,Y) \sim \gamma$, alors $X \sim \mu$ et $Y \sim \nu$; on dira d'ailleurs souvent de manière abusive que (X,Y) est un couplage de μ et ν. Tout l'intérêt des méthodes de couplage réside dans le choix du bon couplage de μ et ν, c'est à dire dans la façon dont X et Y sont inter-dépendantes. Par exemple, $\mu \otimes \nu$ est un couplage de μ et ν, le couplage indépendant, qui n'est pas particulièrement intéressant en règle générale mais qui a le mérite d'assurer l'existence de couplages. Armés de la notion de couplage, on peut donner une autre caractérisation des distances mentionnées plus haut.

Proposition 1.2.1 (*Dualité*)

Soient $\mu, \nu \in \mathscr{M}_1$, et f, g leurs densités respectives par rapport à une mesure λ. On a

$$\|\mu - \nu\|_{TV} = \inf_{X\sim\mu, Y\sim\nu} \mathbb{P}(X \neq Y) = 1 - \int (f \wedge g) d\lambda = \frac{1}{2}\int |f-g| d\lambda. \tag{1.2.3}$$

Si de plus $\int |x|\mu(dx) < +\infty$ et $\int |x|\nu(dx) < +\infty$, alors on a

$$W_1(\mu,\nu) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[|X-Y|]. \tag{1.2.4}$$

Mentionnons au passage que $\mu \ll \mu + \nu$ et $\nu \ll \mu + \nu$, on dispose donc d'un choix facile pour λ. D'autres possibilités naturelles sont les mesures de Lebesgue ou de comptage, suivant le cadre du problème. Notons au passage que la définition à l'aide d'un infimum est la bienvenue lorsqu'il s'agit de majorer une distance, ce qui est nécessaire en pratique, puisqu'un seul couplage fournit une majoration de la distance souhaitée. À nous de trouver le meilleur couplage possible. Montrer que (1.2.2) et (1.2.4) sont équivalentes est difficile, il s'agit du théorème de Kantorovitch-Rubinstein qu'on ne démontrera pas ici. On peut par contre démontrer l'équivalence entre (1.2.1) et (1.2.3) plus facilement, et cette preuve a l'avantage d'exhiber le couplage optimal en variation totale, c'est-à-dire le choix de X et Y qui minimise $\mathbb{P}(X \neq Y)$; on pourra consulter à ce sujet [Lin92]. Avant de prouver ce résultat, soulignons par un exemple un fait important : la distance en variation totale est très qualitative, alors que la distance de Wasserstein est plutôt quantitative. En effet, s'il suffit pour deux variables aléatoires d'être proches l'une de l'autre pour avoir une petite distance de Wasserstein, il leur faut être égales pour avoir une petite distance en variation totale :

$$\|\delta_x - \delta_y\|_{TV} = \mathbb{1}_{x\neq y}, \quad W_1(\delta_x, \delta_y) = |x-y|. \tag{1.2.5}$$

Démonstration de la Proposition 1.2.1 : Comme indiqué plus haut, on ne va démontrer que (1.2.3). Pour la démonstration de 1.2.4, on pourra consulter [dA82, Appendice B]. Notons $A^\star = \{f \leq g\}$ et $p = \int (f \wedge g) d\lambda$, et commençons par remarquer que, puisque μ et ν sont des mesures de probabilité,

$$1 - \int (f \wedge g) d\lambda = \frac{1}{2}\int |f-g| d\lambda = 1 - p.$$

Le calcul est rapide, mais l'intérêt réside plutôt dans un schéma (voir Figure 1.2.1).

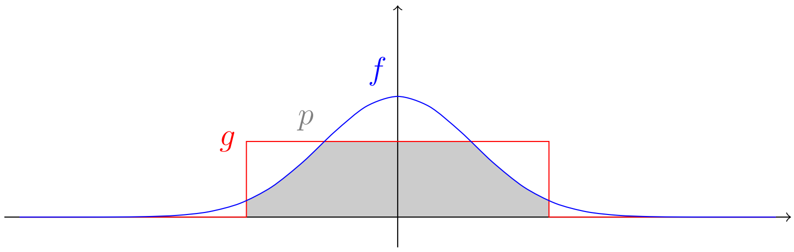


FIGURE 1.2.1 – Distance en variation totale entre $\mathscr{N}(0,1)$ et $\mathscr{U}([-2,2])$; $\|\mathscr{N}(0,1) - \mathscr{U}([-2,2])\|_{TV} = 1-p$.

Ensuite, remarquons que

$$|\mu(A^\star) - \nu(A^\star)| = \int_{A^\star} (g-f)d\lambda = 1-p,$$

donc $1-p \leq \|\mu-\nu\|_{TV}$. Maintenant, pour tous $X \sim \mu, Y \sim \nu$ et $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}|\mu(A) - \nu(A)| = |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| &= |\mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y)| \\ &\leq \mathbb{P}(X \neq Y),\end{aligned}$$

d'où $\|\mu-\nu\|_{TV} \leq \inf_{X\sim\mu,Y\sim\nu} \mathbb{P}(X \neq Y)$. Il ne reste plus qu'à exhiber un couplage tel que $\mathbb{P}(X = Y) \geq p$. Pour cela, on définit $B \sim \mathscr{B}(p)$ et

- si $B = 1$, on pose $X \sim \frac{1}{p}(f \wedge g)\lambda$ et $Y = X$.
- si $B = 0$, on pose $X \sim \frac{1}{1-p}(f - f \wedge g)\lambda$ et $Y \sim \frac{1}{1-p}(g - f \wedge g)\lambda$.

Si $B = 1$, $X = Y$ donc $\mathbb{P}(X = Y) \geq p$. Il reste à vérifier que (X, Y) est un couplage de μ et ν c'est-à-dire que $X \sim \mu$ et $Y \sim \nu$. On a, pour tout borélien A,

$$\begin{aligned}\mathbb{P}(X \in A) &= \mathbb{P}(X \in A, B = 1) + \mathbb{P}(X \in A, B = 0) \\ &= p\int_A \frac{1}{p}(f \wedge g)d\lambda + (1-p)\int_A \frac{1}{1-p}(f - f\wedge g)d\lambda = \int_A f d\lambda = \mu(A).\end{aligned}$$

De même, $\mathbb{P}(Y \in A) = \nu(A)$. □

Remarque 1.2.2 (*Couplage optimal pour* W_1) : Nous avons parlé du couplage optimal pour la distance en variation totale, mais qu'en est-il du couplage optimal pour la distance de Wasserstein ? Tout d'abord, il n'y a a priori pas unicité du couplage optimal : par exemple, nous n'avons pas choisi l'inter-dépendance entre X et Y si $B = 0$ dans le cas du couplage fourni dans la preuve de la Proposition 1.2.1. Pour ce qui est de l'existence (l'infimum est-il atteint ?), ce n'est pas toujours évident, et le lecteur intéressé pourra consulter [AGS08, Théorème 6.2.4] ou [Vil09, Théorème 5.9]. À titre d'exemple, on se contentera de donner un couplage optimal pour W_1 en dimension 1, appelé *réarrangement croissant*. On suppose donc que μ et ν sont des probabilités sur $\mathbb{R}$, dont les fonctions de répartition respectives admettent pour inverse généralisé F^{-1} et G^{-1}. Alors, si $U \sim \mathscr{U}([0,1])$, on définit

$$X = F^{-1}(U), \quad Y = G^{-1}(U),$$

et $W_1(\mu,\nu) = \mathbb{E}[|X-Y|]$. ◇

Enfin, concluons cette section en évoquant un autre type de distance sur l'espace des lois de probabilité. Si $\mathscr{F}$ est une classe de fonctions, on définira

$$d_{\mathscr{F}}(\mu,\nu) = \sup_{\varphi\in\mathscr{F}} |\mu(\varphi)-\nu(\varphi)|.$$

Par exemple, si $\mathscr{F} = \mathscr{C}_b^1$, $d_{\mathscr{F}}$ est une distance appelée distance de Fortet-Mourier, et est connue pour métriser la convergence en loi. En règle générale, $d_{\mathscr{F}}$ est une pseudo-distance, mais il s'agit d'une distance dès que $\mathscr{F}$ contient une algèbre de fonctions continues bornées qui sépare les points (voir [EK86, Théorème 4.5.(a), Chapitre 3]). Dans tous les cas traités dans ce manuscrit, $\mathscr{F}$ contient l'algèbre $\mathscr{C}_c^\infty$ "à constante près", et donc la convergence au sens de $d_{\mathscr{F}}$ entraîne la convergence en loi, comme le souligne le résultat suivant (qui sera prouvé au Chapitre 4).

Lemme 1.2.3 (*Convergence en loi et $d_{\mathscr{F}}$*)

Soient $(\mu_n),\mu$ des mesures de probabilité. Supposons que $\mathscr{F}$ soit étoilé par rapport à 0 (i.e. si $\varphi\in\mathscr{F}$ alors $\lambda\varphi\in\mathscr{F}$ pour $\lambda\in[0,1]$) et que, pour tout $\psi\in\mathscr{C}_c^\infty$, il existe $\lambda>0$ tel que $\lambda\psi\in\mathscr{F}$. Si $\lim_{n\to\infty} d_{\mathscr{F}}(\mu_n,\mu)=0$, alors (μ_n) converge en loi vers μ. Si de plus $\mathscr{F}\subseteq\mathscr{C}_b^1$, alors $d_{\mathscr{F}}$ métrise la convergence en loi.

Il est à noter que ce cadre capture les distances en variation totale et de Wasserstein introduites auparavant. En particulier, le Lemme 1.2.3 permet de voir que les convergences au sens de ces distances sont strictement plus fortes que la convergence en loi :

- la convergence en W_1 est classiquement équivalente à la convergence en loi adjointe à la convergence du premier moment.
- Dans $\mathbb{R}$ muni de sa topologie usuelle, $(\delta_{1/t})_{t\geq 0}$ converge en loi vers δ_0 mais $\|\delta_{1/t}-\delta_0\|_{TV}=1$, car leurs lois sont à supports disjoints. Par contre, de manière générale, la convergence en variation totale est équivalente à la convergence en loi dans un espace de probabilité fini ou dénombrable.

1.2.2 Ergodicité exponentielle

Dans cette section, nous allons voir comment l'on peut quantifier la vitesse de convergence d'un processus de Markov $(X_t)_{t\geq 0}$ vers sa mesure stationnaire π, c'est-à-dire quantifier $W_1(\mathscr{L}(X_t),\pi)$ ou $\|\mathscr{L}(X_t)-\pi\|_{TV}$. On parlera d'*ergodicité exponentielle* lorsque ces quantités sont majorées par une vitesse $C\mathrm{e}^{-vt}$, avec $C,v>0$.

La première méthode que nous aborderons est le critère de Foster-Lyapunov, qui est notamment exposé de manière exhaustive dans [MT93a] (citons aussi les article plus accessibles [MT93b, DMT95]) ; il est d'ailleurs souvent fait référence à ces idées comme *techniques à la Meyn et Tweedie*. Notons $\mathcal{L}$ le générateur infinitésimal de (X_t) et, pour $t\geq 0$, $\mu_t=\mathscr{L}(X_t)$. L'idée est de trouver une fonction V, dite de Lyapunov, contrôlant les excursions de (X_t) hors d'un compact. On dira d'un ensemble K qu'il est petit[4] pour $(X_t)_{t\geq 0}$ s'il existe une mesure de probabilité $\mathscr{A}$ sur $\mathbb{R}_+$ et une mesure positive non-triviale ν sur $\mathbb{R}^d$ telles que, pour tout $x \in K$, $\int_0^\infty \delta_x P_t \mathscr{A}(dt) \geq \nu$. On donnera une interprétation de cette notion à la Remarque 1.2.7 ; pour le moment, nous donnons le fameux critère.

Théorème 1.2.4 (*Critère de Foster-Lyapunov*)

Soient V une fonction coercive strictement positive, $K \subseteq \mathbb{R}^d$ petit pour $(X_t)_{t\geq 0}$ et $\alpha, \beta > 0$. Si X est irréductible et apériodique (voir [DMT95]), si V est bornée sur K et si

$$\mathcal{L}V(x) \leq -\alpha V(x) + \beta \mathbb{1}_K(x), \tag{1.2.6}$$

alors $(X_t)_{t\geq 0}$ possède une unique mesure stationnaire π telle que $\pi(V) < +\infty$, et il existe $C, v > 0$ tels que

$$d_{\mathscr{F}}(\mu_t, \pi) \leq C\mu(V)e^{-vt},$$

où $\mathscr{F} = \{\varphi \in \mathscr{C}^0 : |\varphi| \leq V + 1\}$. En particulier,

$$\|\mu_t - \pi\|_{TV} \leq C\mu(V)e^{-vt}.$$

Nous verrons des exemples d'application de ce théorème au Chapitre 3. La convergence en variation totale est assurée par l'inclusion $\{\varphi \in \mathscr{C}^0 : \|\varphi\|_\infty \leq 1\} \subseteq \mathscr{F}$. Le Théorème 1.2.4 est très général et très puissant : il fournit en effet l'existence et l'unicité de π ainsi qu'une vitesse de convergence vers celle-ci dans une distance plus forte que la variation totale. Par contre, on lui reprochera de ne pas donner explicitement les constantes C et v, ce qui en fait un résultat somme toute très théorique. Signalons qu'il reste possible de suivre les démonstrations pour obtenir des constantes explicites, qui sont alors généralement très mauvaises par rapport à ce qu'on pourrait obtenir avec d'autres méthodes. Il n'empêche qu'il s'agit d'une méthode très utilisée en pratique. Il existe d'ailleurs de nombreux critères similaires, permettant de caractériser différentes propriétés du processus de Markov (non-explosion, transience, récurrence, positivité...). Terminons cette description du critère de Foster-Lyapunov en signalant que la littérature abonde d'autres versions et raffinements de ce résultat, qui traitent par exemple des chaînes de Markov inhomogènes ou de vitesses de convergence sous-géométrique à l'aide de méthodes variées (on pourra consulter par exemple [DMR04, DFG09, HM11]). La construction de fonction de Lyapunov pour un PDMP est en général assez aisée, et le lecteur intéressé pourra trouver des idées dans le Chapitre 3, ainsi que dans les articles [BCT08, MH10].

Remarque 1.2.5 (*Condition suffisante pour être une fonction de Lyapunov*) : On notera qu'une condition suffisante pour qu'une fonction V continue vérifie (1.2.6) est l'existence d'une fonction f continue sur $\mathbb{R}^d$, telle que

$$\mathcal{L}V(x) \leq f(x)V(x), \qquad \lim_{|x|\to+\infty} f(x) = -\infty.$$

[4] On parle de *petite set* en anglais, qui est différent d'un *small set*.

En effet, il existe $A > 0$ tel que, en notant $K = \bar{B}(0, A)$, $f \leq -1$ sur K^C. Alors

$$\mathcal{L}V \leq -V + \sup_K((f+1)V)\mathbb{1}_K.$$

$\diamondsuit$

Nous adoptons maintenant un autre point de vue, en cherchant à quantifier la vitesse de convergence exponentielle obtenue plus haut ; nous allons faire appel à des méthodes de couplage, et justifier l'existence de la Section 1.2.1. L'idée est de construire intelligemment un couplage $(X, \widetilde{X})$ constitué de deux processus de Markov suivant chacun la dynamique dictée par $\mathcal{L}$, ce qui revient à construire un processus de Markov dans $\mathbb{R}^{2d}$, et tel que $\lim_{t\to\infty} d(\mu_t, \widetilde{\mu}_t) = 0$ (où l'on a noté $\mu_t = \mathscr{L}(X_t)$, $\widetilde{\mu}_t = \mathscr{L}(\widetilde{X}_t)$ et d une certaine distance sur $\mathscr{M}_1$). En effet, si $\widetilde{\mu}_0 = \pi$, alors pour tout $t \geq 0$, $\mu_t = \pi$ et $\lim_{t\to\infty} d(\mu_t, \pi) = 0$. On peut alors estimer cette vitesse de convergence dans la distance qui nous intéresse. Une variable aléatoire essentielle dans cette étude est l'instant de couplage des deux processus :

$$\tau = \inf\{t \geq 0 : \forall s \geq 0, X_{t+s} = \widetilde{X}_{t+s}\}.$$

On notera un certain flou concernant le terme *couplage*, qui désigne à la fois une loi sur l'espace produit, un couple suivant cette loi, et le fait que deux versions d'un processus deviennent égales (notons aussi l'usage du terme *coalescence* dans ce cas). Notons que l'instant de couplage n'est a priori pas un temps d'arrêt par rapport à la filtration engendrée par $(X, \widetilde{X})$, mais il est généralement possible de s'en assurer avec une bonne construction du couplage, puisque l'on est dans un cadre markovien. Ensuite, en notant ψ_τ la transformée de Laplace de τ, il est facile de voir que

$$\|\mu_t - \widetilde{\mu}_t\|_{TV} \leq \mathbb{P}(X_t \neq \widetilde{X}_t) \leq \mathbb{P}(\tau > t) \leq \psi_\tau(u)\mathrm{e}^{-ut}, \tag{1.2.7}$$

dès que τ admet un moment exponentiel d'ordre u, c'est-à-dire $\mathbb{E}[\mathrm{e}^{u\tau}] < +\infty$. Plus les trajectoires de X et $\widetilde{X}$ se couplent vite (dans le sens où le temps de couplage est petit), plus la vitesse de convergence à l'équilibre sera rapide. Une excellente référence sur le couplage en variation totale est [Lin92]. Si l'on souhaite obtenir une convergence en Wasserstein, il "suffira" de rapprocher les deux trajectoires sans obligatoirement les rendre égales (rappelons-nous de (1.2.5)).

Remarque 1.2.6 (***Convergence à l'équilibre pour le processus pharmacocinétique***) **:** Nous allons étudier brièvement la vitesse de convergence à l'équilibre du processus pharmacocinétique introduit à la Remarque 1.1.1. Rappelons que, pour $\alpha, \theta, \lambda > 0$,

$$\mathcal{L}f(x) = -\theta x f'(x) + \lambda \int_0^\infty [f(x+u) - f(x)]\alpha \mathrm{e}^{-\alpha u} du.$$

Nous montrerons à la Proposition 3.3.4 que ce processus admet une unique mesure invariante $\pi = \Gamma(\lambda/\theta, 1/\alpha)$. Dans une optique de couplage en Wasserstein, on cherche à choisir de manière conjointe l'aléa dans deux trajectoires de notre PDMP de manière à les rapprocher. Dans notre cas, le flot contracte exponentiellement vite, ce qui est idéal. Les sauts pourraient poser problème, c'est-à-dire éloigner les trajectoires, mais on va pouvoir choisir de les faire sauter au même instant et selon la même amplitude à l'aide d'un couplage *synchrone*. Prenons donc le processus de Markov $(X, \widetilde{X})$ généré par

$$\mathcal{L}_2 f(x, \widetilde{x}) = -\theta \partial_x f(x, \widetilde{x}) - \theta \partial_{\widetilde{x}} f(x, \widetilde{x}) + \lambda \int_0^\infty [f(x+u, \widetilde{x}+u) - f(x, \widetilde{x})] \alpha \mathrm{e}^{-\alpha u} du.$$

Remarquons que, si $f(x, \widetilde{x}) = f_1(x)$ ou $f_2(\widetilde{x})$, on vérifie aisément que $\mathcal{L}_2$ coïncide avec $\mathcal{L}$, ce qui signifie que les processus X et $\widetilde{X}$ pris séparément suivent la dynamique attendue. Mais qu'arrive-t-il au couple ? Le terme de dérive assure une décroissance exponentielle de chaque trajectoire à taux θ et, à des instants séparés par une variable aléatoire de loi $\mathscr{E}(\lambda)$, les deux processus sautent en même temps vers le haut suivant une même variable aléatoire de loi $\mathscr{E}(\alpha)$. Le point important est que le saut est le même pour chaque processus, et ne se voit donc pas lorsque l'on regarde leur écart. Cette dynamique est illustrée à la Figure 1.2.2. Pour commencer, supposons que $X_0 = x \geq \widetilde{X}_0 = \widetilde{x}$. Le processus X reste alors toujours supérieur à $\widetilde{X}$ (on parle de *couplage monotone*) et on a

$$W_1(\mu_t, \widetilde{\mu}_t) \leq \mathbb{E}[|X_t - \widetilde{X}_t|] = (x - \widetilde{x})\mathrm{e}^{-\theta t}.$$

Maintenant, si μ_0 et $\widetilde{\mu}_0$ sont deux lois quelconques, choisissons $(X_0, \widetilde{X}_0)$ comme le couplage optimal de μ_0 et $\widetilde{\mu}_0$ en W_1 comme défini à la Remarque 1.2.2. On obtient alors

$$W_1(\mu_t, \widetilde{\mu}_t) \leq W_1(\mu_0, \widetilde{\mu}_0)\mathrm{e}^{-\theta t}.$$

On obtient une contraction en distance de Wasserstein, ce qui est généralement difficile à obtenir mais peut être très utile. D'après les simulations (voir Figure 1.2.3), cette majoration donne la vraie vitesse de convergence en W_1. Dans certains cas simples, la vitesse de décroissance en Wasserstein est non seulement majorable, mais directement calculable grâce à la notion de courbure de Wasserstein (voir par exemple [Jou07, Clo13]), mais nous n'en parlerons pas plus ici.

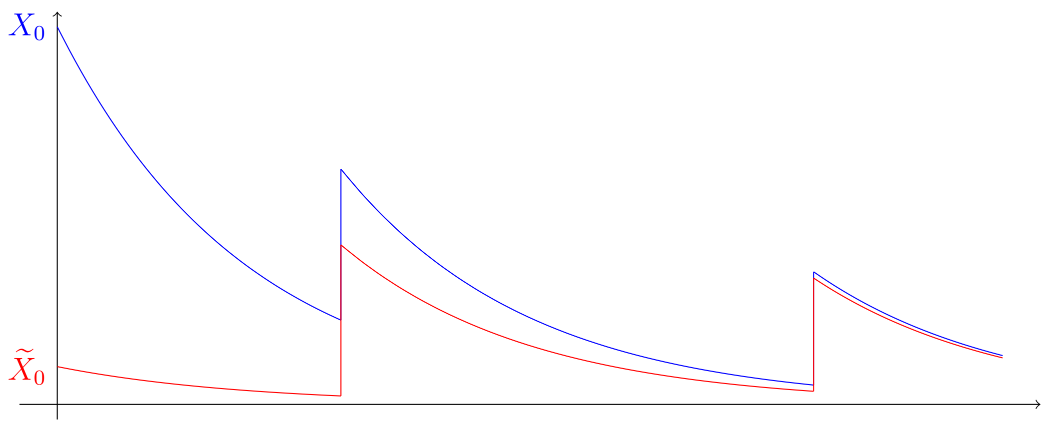


FIGURE 1.2.2 – Comportement typique du couplage défini à la Remarque 1.2.6.

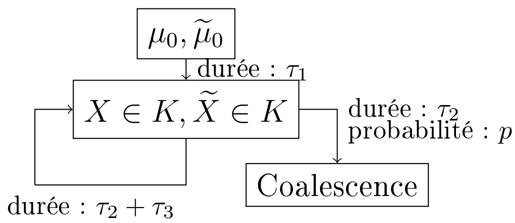
$\diamondsuit$

Notons que l'on pourrait obtenir d'une manière proche une convergence en variation totale, ce qui sera traité dans un cadre plus général au Chapitre 2. Si l'on veut donner brièvement l'heuristique, il s'agit de rapprocher les deux processus grâce au couplage monotone utilisé plus haut, puis de les faire sauter au même endroit en s'appuyant sur la densité de la loi du saut $\mathscr{E}(\alpha)$.

FIGURE 1.2.3 – Tracé de $W_1(\mu_t, \pi)$ en fonction de t, pour $\mu_0 = \delta_5, \theta = 1, \lambda = 0.5, \alpha = 2$.

Concluons ce tour d'horizon du couplage en citant quelques articles traitant de ces méthodes de couplage, que ce soit en Wasserstein ou en variation totale. Par exemple, [CMP10, BCG+13b] ont introduit dans le cadre du processus TCP les méthodes utilisées dans ce manuscrit, et plus particulièrement dans le Chapitre 2. L'article [BCF15] traite de méthodes de couplage pour les processus de renouvellement, d'une manière différente de celle que nous verrons au Chapitre 2. On trouve aussi des méthodes similaires dans [FGM12, FGM15] concernant les processus de télégraphe.

Remarque 1.2.7 (*Foster-Lyapunov vu comme un couplage*) : Il est intéressant de remarquer que les hypothèses du Théorème 1.2.4 peuvent s'interpréter comme des conditions pour obtenir une convergence en variation totale à l'aide de méthodes de couplage. En effet, on demande au processus de Markov $(X_t)_{t\geq 0}$ d'admettre une fonction de Lyapunov (inégalité (1.2.6)) et aux ensembles compacts d'être petits. On peut alors créer un couplage $(X_t, \widetilde{X}_t)$ dont l'heuristique est la suivante :



- En partant de l'état initial $(\mu_0, \widetilde{\mu}_0)$, on amène X et $\widetilde{X}$ dans l'ensemble petit K (typiquement, un compact) en une durée τ_1.
- Avec une probabilité au moins égale à p, on amène à coalescence X et $\widetilde{X}$ en un

temps τ_2. La probabilité p est uniforme en les points de départ des deux processus à l'intérieur de K. Ce mécanisme utilise le fait que K soit petit, au sens du Théorème 1.2.4. La mesure ν permet de quantifier la probabilité de couplage, au bout d'un temps suivant une loi $\mathscr{A}$.

- Si X et $\widetilde{X}$ n'ont pas été couplés, on attend un temps τ_3 nécessaire pour que X et $\widetilde{X}$ reviennent dans K, puis on réessaie de les coupler. Il est nécessaire de contrôler τ_3, et cela se fait à l'aide de la fonction de Lyapunov.

Mettre en place une telle dynamique n'est pas particulièrement évident (on consultera plutôt [MT93a] pour les détails) et l'on ne s'y aventurera pas ici. Néanmoins, quand cela fonctionne, le temps de couplage des deux processus est égal à

$$\tau = \tau_1 + \tau_2 + G(\tau_2 + \tau_3),$$

où G suit une loi géométrique $\mathscr{G}(p)$ (le nombre d'essais ratés). Il est alors possible de montrer que τ admet des moments exponentiels, ce qui implique l'ergodicité exponentielle d'après (1.2.7). $\diamondsuit$

Outre les méthodes de Foster-Lyapunov et de couplage, citons une autre grande famille de techniques à caractère très analytique : les inégalités fonctionnelles. On pourra consulter à ce sujet [Bak94, ABC+00, BCG08, Mon14b]. L'idée est d'obtenir des inégalités fonctionnelles mettant en jeu la mesure invariante π et le générateur infinitésimal $\mathcal{L}$ du processus concerné. Par exemple, en notant

$$\Gamma f = \frac{1}{2}\mathcal{L}f^2 - f\mathcal{L}f,$$

l'opérateur *carré du champ*, on remarque que $\Gamma f \geq 0$ et que, par invariance, $\mu(\Gamma f) = -\mu(f\mathcal{L}f)$. On dit que π vérifie une inégalité de Poincaré de constante C si, pour toute fonction régulière f,

$$\mathrm{Var}_\pi(f) = \pi(f^2) - \pi(f)^2 \leq C\pi(\Gamma f).$$

On peut alors montrer le théorème suivant, reliant l'inégalité de Poincaré à l'ergodicité exponentielle, et faisant intervenir de manière un peu technique une algèbre $\mathcal{A}$ de fonctions définie par exemple dans [ABC+00, Définition 2.4.2].

Théorème 1.2.8 (*Inégalité de trou spectral*)

Les deux assertions suivantes sont équivalentes :

i) π vérifie une inégalité de Poincaré de constante C.

ii) Pour toute fonction $f \in \mathcal{A}$,

$$\|P_t f - \pi(f)\|_{L^2(\pi)} \leq \sqrt{\mathrm{Var}_\pi(f)}e^{-\frac{1}{C}t}.$$

Le Théorème 1.2.8 est qualifié d'*inégalité de trou spectral* car la constante optimale $1/C$ correspond au trou spectral de l'opérateur $\mathcal{L}$, c'est-à-dire à l'opposé de la première valeur propre non-nulle (quand elle existe) de $\mathcal{L}$. Celui-ci n'admet en effet que des valeurs propres de parties réelles négatives, ainsi que 0 associé aux constantes. Cela se démontre en effectuant une décomposition spectrale de $P_t f$; on pourra trouver plus de détails dans [Bak94]. En tout cas, il s'agit d'une manière de faire le lien entre analyse spectrale et inégalités fonctionnelles. D'autres inégalités fonctionnelles existent, parmi lesquelles les inégalités de Sobolev logarithmiques (ou log-Sobolev), lorsqu'on travaille avec l'entropie plutôt qu'avec la variance, et qui sont strictement plus fortes que les inégalités de Poincaré.

Il est possible, comme dans [BCG08, CGZ13], de faire la correspondance (parfois même quantitative) entre l'inégalité de Poincaré, le critère de Foster-Lyapunov et la convergence exponentielle à l'équilibre dans le cas de certains processus réversibles. En revanche, si les processus ne sont pas réversibles, comme c'est le cas pour les PDMP que nous étudierons dans la suite de ce manuscrit, les choses ne se passent pas aussi bien. On citera quand même l'article [Mon15] qui adapte des critères classiques d'inégalités fonctionnelles au cas de certains PDMP en obtenant des inégalités fonctionnelles pour un autre carré-du-champ que celui associé à $\mathcal{L}$.

Remarque 1.2.9 (*Ergodicité du processus d'Ornstein-Uhlenbeck*) : Illustrons sur un exemple-type le lien entre ces différentes méthodes quantifiant la vitesse de convergence à l'équilibre d'un processus de Markov : le processus Ornstein-Uhlenbeck sur $\mathbb{R}$. À noter que les résultats de cette remarque s'étendent facilement au processus d'Ornstein-Uhlenbeck sur $\mathbb{R}^d$. Ce processus n'est pas un PDMP, mais un processus diffusif, qui satisfait l'EDS suivante

$$dX_t = -X_t + \sqrt{2} dW_t, \quad X_0 \sim \mu,$$

où W est un mouvement brownien. Alternativement, on peut le définir par son générateur infinitésimal

$$\mathcal{L}f(x) = -xf'(x) + f''(x).$$

Une vérification directe par intégration par parties nous assure que la mesure de probabilité invariante associée à $(X_t)_{t\geq 0}$ est $\pi = \mathcal{N}(0,1)$. Tout d'abord, vérifions que $V(x) = \exp(\theta|x|)$ est une fonction de Lyapunov pour X pour tout $\theta > 0$. On a

$$\mathcal{L}V(x) = \left(-\theta|x| + \theta^2\right) V(x), \quad \lim_{x\to\pm\infty} -\theta|x| + \theta^2 = -\infty.$$

La fonction V satisfait donc (1.2.6) en vertu de la Remarque 1.2.5, et les autres hypothèses du Théorème 1.2.4 sont satisfaites, si bien que la loi de X converge exponentiellement vers $\pi = \mathcal{N}(0,1)$.

La loi normale centrée réduite vérifie une inégalité de Poincaré de constante optimale $C = 1$ (voir par exemple [ABC+00, Théorème 1.5.1]), et le Théorème 1.2.8 nous assure donc que, pour toute fonction $f \in \mathcal{A}$,

$$\|P_t f - \pi(f)\|_{L^2(\pi)} \leq \sqrt{\mathrm{Var}_\pi(f)} \mathrm{e}^{-t}. \tag{1.2.8}$$

D'autre part, X s'obtient explicitement en fonction de W par la formule suivante, aisément vérifiable à l'aide de la formule d'Itô :

$$X_t = X_0 \mathrm{e}^{-t} + \sqrt{2} \int_0^t \mathrm{e}^{-(t-s)} dW_s.$$

Considérons $\widetilde{X}$ un autre processus d'Ornstein-Uhlenbeck de même dynamique, de loi initiale $\widetilde{\mu}$ telle que $(X_0, \widetilde{X}_0)$ soit le couplage optimal en W_1 de μ et $\widetilde{\mu}$ et tel que

$$\widetilde{X}_t = \widetilde{X}_0 \mathrm{e}^{-t} + \sqrt{2} \int_0^t \mathrm{e}^{-(t-s)} dW_s.$$

Le processus W étant le même mouvement brownien dirigeant X et $\widetilde{X}$, on a directement

$$\mathbb{E}[X_t - \widetilde{X}_t] = W_1(\mu, \widetilde{\mu})\mathrm{e}^{-t}.$$

Si $\widetilde{\mu} = \pi$, on a alors

$$W_1(\mathscr{L}(X_t), \pi) = W_1(\mu, \pi)\mathrm{e}^{-t}. \tag{1.2.9}$$

La vitesse de décroissance dans (1.2.8) est la même que dans (1.2.9). Ce n'est pas un résultat général, et une méthode donnera dans certains cas de meilleurs résultats qu'une autre, dépendant fortement de la finesse des estimés des méthodes de couplage ou des inégalités mises en jeu lors du calcul de la constante de Poincaré. Cette dernière méthode tombera généralement en défaut si le processus n'est pas réversible. ◇

1.2.3 Pour aller plus loin

Pour renforcer les résultats énoncés dans les sections précédentes, on peut s'intéresser à la loi de $(X_t)_{t\geq 0}$ en tant que processus, et non pas à la loi de X_t pour t fixé. Le cadre naturel de cette section est donc l'espace de Skorokhod des fonction càdlàg, puisque tout processus de Markov admet une version càdlàg p.s. s'il est Feller ; des références classiques sont [Bil99, JS03]. Il est possible de munir l'espace de Skorokhod d'une métrique qui en fait un espace polonais, et qui coïncide avec celle de la convergence uniforme sur tout compact lorsqu'on se restreint à l'espace des fonctions continues ; voir [JM86] par exemple.

La convergence de lois de probabilité sur l'espace de Skorokhod est généralement appelée *convergence fonctionnelle*, et s'obtient de manière classique en prouvant la tension[5] de la suite de mesures, adjointe à la convergence des lois fini-dimensionnelles. La tension assure la relative compacité de la suite, tandis que les lois fini-dimensionnelles caractérisent la limite obtenue. Cette architecture de preuve sera par exemple utilisée au Chapitre 4 pour prouver la convergence en loi du processus interpolé vers un processus limite sur un intervalle de temps $[0, T]$. Un critère classique de tension est le critère d'Aldous-Rebolledo qu'on trouvera par exemple énoncé dans [JM86, Théorème 2.2.2 et 2.3.2].

Il n'est parfois pas possible d'étudier directement la convergence d'une famille de mesures de probabilité $(\mu_t)_{t\geq 0}$ vers une certaine loi π. Dans certains cas, on pourra passer par l'intermédiaire d'un processus de Markov dont la loi au temps t est "proche" de μ_t, et qui est ergodique de mesure stationnaire π. C'est le problème soulevé au Chapitre 4. Nous définissons donc la notion de pseudo-trajectoire asymptotique, introduite

[5] On parle de *tightness* en anglais.

dans [BH96] (on pourra aussi consulter [Ben99]). Grâce à la relation de Chapman-Kolmogorov, on peut voir le semi-groupe (P_t) d'un processus de Markov (X_t) comme un semi-flot sur l'espace des mesures de probabilité, que l'on note

$$\Phi(\mu, t) = \mu P_t.$$

Considérons une famille de mesures de probabilité $(\mu_t)_{t\geq 0}$ et une distance d sur $\mathscr{M}_1$. On dit que (μ_t) est une pseudo-trajectoire asymptotique de Φ par rapport à d si, pour tout $T > 0$,

$$\lim_{t\to\infty} \sup_{0\leq s\leq T} d(\mu_{t+s}, \Phi(\mu_t, s)) = 0.$$

De même, on dira que (μ_t) est une λ-pseudo-trajectoire de Φ (par rapport à d) s'il existe $\lambda > 0$ tel que, pour tout $T > 0$,

$$\limsup_{t\to+\infty} \frac{1}{t} \log \left(\sup_{0\leq s\leq T} d(\mu_{t+s}, \Phi(\mu_t, s)) \right) \leq -\lambda.$$

La notion de λ-pseudo-trajectoire permet de quantifier celle de pseudo-trajectoire asymptotique et, si X est exponentiellement ergodique, permet d'obtenir des vitesses de convergences similaires pour (μ_t).

1.2.4 Applications de l'ergodicité

Il existe un lien très fort entre les processus de Markov et certaines équations aux dérivées partielles. En effet, si la loi d'un processus de Markov à l'instant t admet une densité, celle-ci vérifie une Equation aux Dérivées Partielles (EDP) intrinsèquement liée à la dynamique du processus. Si X est un processus de Markov de semi-groupe (P_t) et de générateur infinitésimal $\mathcal{L}$, nous avons vu à la Section 1.1.1 que

$$\partial_t(P_t f) = \mathcal{L} P_t f$$

Il est rapide de vérifier qu'il s'agit de la formulation faible de

$$\partial_t \mu_t = \mathcal{L}' \mu_t, \tag{1.2.10}$$

où $\mu_t = \mathscr{L}(X_t)$ et $\mathcal{L}'$ est l'opérateur adjoint naturel de $\mathcal{L}$, au sens L^2. On réservera la notation $\mathcal{L}^*$ au générateur des processus retournés en temps que l'on introduira au Chapitre 3, qui est l'adjoint de $\mathcal{L}$ dans $L^2(\pi)$. Dans le cadre d'un processus diffusif, l'équation (1.2.10) est appelée *équation de Fokker-Planck*. L'étude en temps long du processus de Markov ou celle de l'EDP vérifiée par sa densité sont des problèmes aux thématiques proches mais dont les outils de résolution sont assez différents. Soulignons que les inégalités fonctionnelles sont l'un des outils à l'intersection des deux domaines (voir par exemple [AMTU01, Gen03]). Nous verrons à la Section 3.2.1 comment l'on peut étudier une EDP du type de (1.2.10) avec des outils probabilistes, en ayant besoin d'hypothèses similaires pour que tout se passe bien.

Les statistiques sont aussi un domaine dans lequel la compréhension du comportement en temps long d'un processus de Markov est très importante. Obtenir des bornes fines sur les vitesses de convergence à l'équilibre est crucial pour pouvoir mettre en place des modèles statistique efficaces, par exemple pour estimer le temps passé au-dessus de certains seuils de dangerosité dans le cadre de modèles de pharmacocinétique. En effet, il est courant en statistiques de considérer que des processus sont à l'équilibre après un "certain temps", et la question de spécifier précisément ce "certain temps" se pose naturellement. Dans le cadre de la pharmacocinétique, on pourra consulter [GP82] pour les motivations et [CT09, BCT10] pour les applications de l'ergodicité aux statistiques. À noter que ces seuils reçoivent beaucoup d'attention dans le domaines des processus de type shot-noise (voir par exemple [OB83, BD12]), et que l'on peut sous certaines hypothèses établir une correspondance entre shoit-noise et PDMP, comme on le verra au Chapitre 3. Récemment, l'estimation des paramètres des PDMP a aussi suscité beaucoup d'attention de la part de la communauté mathématique. Une question très actuelle est l'estimation du taux de saut, et de savoir de quoi celui-ci dépend ; citons par exemple [DHRBR12, RHK+14, DHKR15] dans le cadre des modèles de croissance/fragmentation, ou [ADGP14, AM15] dans un cadre plus général. Là encore, la compréhension des mécanismes du PDMP est cruciale pour mettre en place des modèles fins.

CHAPTER 2

PIECEWISE DETERMINISTIC MARKOV PROCESSES AS A MODEL OF DIETARY RISK

In this chapter, we consider a *Piecewise Deterministic Markov Process* (PDMP) modeling the quantity of a given food contaminant in the body. On the one hand, the amount of contaminant increases with random food intakes and, on the other hand, decreases thanks to the release rate of the body. Our aim is to provide quantitative speeds of convergence to equilibrium for the total variation and Wasserstein distances via coupling methods.

Note: this chapter is an adaptation of [Bou15].

2.1 Introduction

We study a PDMP modeling pharmacokinetic dynamics; we refer to [BCT08] and the references therein for details on the medical background motivating this model. This process is used to model the exposure to some chemical, such as methylmercury, which can be found in food. It has three random parts: the amount of contaminant ingested, the inter-intake times and the release rate of the body. Under some simple assumptions, with the help of Foster-Lyapounov methods, the geometric ergodicity has been proven in [BCT08]; however, the rates of convergence are not explicit. The goal of our present paper is to provide quantitative exponential speeds of convergence to equilibrium for this PDMP, with the help of coupling methods. Note that another approach, quite recent, consists in using functional inequalities and hypocoercive methods (see [Mon14a, Mon15]) to quantify the ergodicity of non-reversible PDMPs.

Firstly, let us present the PDMP introduced in [BCT08], and recall its infinitesimal generator. We consider a test subject whose blood composition is constantly monitored. When he eats, a small amount of a given food contaminant (one may think of methylmercury for instance) is ingested; denote by X_t the quantity of the contaminant in the body at time t. Between two contaminant intakes, the body purges itself so that the process X follows the ordinary differential equation

$$\partial_t X_t = -\Theta X_t,$$

where $\Theta > 0$ is a random metabolic parameter regulating the elimination speed. Following [BCT08], we will assume that Θ is constant between two food ingestions, which makes the trajectories of X deterministic between two intakes. We also assume that the rate of intake depends only on the elapsed time since the last intake (which is realistic for a food contaminant present in a large variety of meals). As a matter of fact, [BCT08] firstly deals with a slightly more general case, where $\partial_t X_t = -r(X_t, \Theta)$ and r is a positive function. Our approach is likely to be easily generalizable if r satisfies a condition like

$$r(x, \theta) - r(\tilde{x}, \theta) \geq C\theta(x - \tilde{x}),$$

but in the present paper we focus on the case $r(x, \theta) = \theta x$.

Define $T_0 = 0$ and T_n the instant of n^{th} intake. The random variables $\Delta T_n = T_n - T_{n-1}$, for $n \geq 2$, are assumed to be independent and identically distributed (i.i.d.) and almost surely (a.s.) finite with distribution G. Let ζ be the hazard rate (or failure rate, see [Lin86] or [Bon95] for some reminders about reliability) of G; which means that $G([0, x]) = 1 - \exp\left(-\int_0^x \zeta(u)du\right)$ by definition. In fact, there is no reason for $\Delta T_1 = T_1$ to be distributed according to G, if the test subject has not eaten for a while before the beginning of the experience. Let $N_t = \sum_{n=1}^\infty \mathbb{1}_{\{T_n \leq t\}}$ be the total number of intakes at time t. For $n \geq 1$, let

$$U_n = X_{T_n} - X_{T_n^-}$$

be the contaminant quantity taken at time T_n (since X is a.s. càdlàg, see a typical trajectory in Figure 2.1.1). Let Θ_n be the metabolic parameter between T_{n-1} and T_n. We assume that the random variables $\{\Delta T_n, U_n, \Theta_n\}_{n\geq 1}$ are independent. Finally, we denote by F and H the respective distributions of U_1 and Θ_1. For obvious reasons, we assume also that the expectations of F and H are finite and $H((-\infty, 0]) = 0$.

From now on, we make the following assumptions (only one assumption among (H4a) and (H4b) is required to be fullfiled):

$$F \text{ admits } f \text{ for density w.r.t. Lebesgue measure.} \tag{H1}$$

$$G \text{ admits } g \text{ for density w.r.t. Lebesgue measure.} \tag{H2}$$

$$\zeta \text{ is non-decreasing and non identically null.} \tag{H3}$$

$$\eta \text{ is Hölder on } [0,1], \text{ where } \eta(x) = \frac{1}{2}\int_{\mathbb{R}} |f(u) - f(u-x)|du. \tag{H4a}$$

$$f \text{ is Hölder on } \mathbb{R}_+ \text{ and there exists } p > 2 \text{ such that } \lim_{x\to+\infty} x^p f(x) = 0. \tag{H4b}$$

From a modeling point of view, (H3) is reasonnable, since ζ models the hunger of the patient. Assumptions (H4a) and (H4b) are purely technical, but reasonably mild.

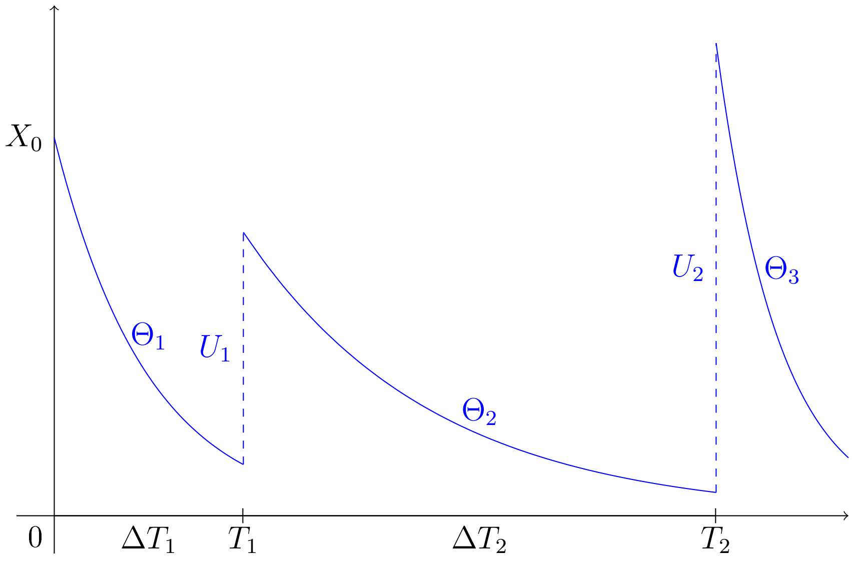


Figure 2.1.1 – Typical trajectory of X.

Note that the process X itself is not Markovian, since the jump rates depends on the time elapsed since the last intake. In order to deal with a PDMP, we consider the process (X, Θ, A), where

$$\Theta_t = \Theta_{N_t+1}, \quad A_t = t - T_{N_t}.$$

We call Θ the metabolic process, and A the age process. The process $Y = (X, \Theta, A)$ is then a PDMP which possesses the strong Markov property (see [Jac06]). Let $(P_t)_{t\geq 0}$ be its semigroup; we denote by $\mu_0 P_t$ the distribution of Y_t when the law of Y_0 is μ_0. Its infinitesimal generator is

$$\begin{aligned} \mathcal{L}\varphi(x,\theta,a) &= \partial_a\varphi(x,\theta,a) - \theta x \partial_x \varphi(x,\theta,a) \\ &+ \zeta(a) \int_0^\infty \int_0^\infty \left[\varphi(x+u,\theta',0) - \varphi(x,\theta,a)\right] H(d\theta')F(du). \end{aligned} \tag{2.1.1}$$

Of course, if ζ is constant, then (X, Θ) is a PDMP all by itself. Let us recall that ζ being constant is equivalent to G being an exponential distribution. Such a model is not relevant in this context, nevertheless it provides explicit speeds of convergence, as it will be seen in Section 2.3.2.

Now, we are able to state the following theorem, which is the main result of our paper; its proof will be postponed to Section 2.3.1.

Theorem 2.1.1

Let $\mu_0, \tilde{\mu}_0$ be distributions on $\mathbb{R}_+^3$. Then, there exist positive constants C_i, v_i (see Remark 2.1.2 for details) such that, for all $0 < \alpha < \beta < 1$:

i) For all $t > 0$,

$$\begin{aligned} \|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq 1 &- \left(1 - C_1 e^{-v_1 \alpha t}\right)\left(1 - C_2 e^{-v_2(\beta-\alpha)t}\right) \\ &\left(1 - C_3 e^{-v_3(1-\beta)t}\right)\left(1 - C_4 e^{-v_4(\beta-\alpha)t}\right). \end{aligned} \tag{2.1.2}$$

ii) For all $t > 0$,

$$W_1(\mu_0 P_t, \tilde{\mu}_0 P_t) \leq C_1 e^{-v_1 \alpha t} + C_2 e^{-v_2(1-\alpha)t}. \tag{2.1.3}$$

Remark 2.1.2: The constants C_i are not always explicit, since they are strongly linked to the Laplace transforms of the distributions considered, which are not always easy to deal with; the reader can find the details in the proof. However, the parameters v_i are explicit and are provided throughout this paper. The speed v_1 comes from Theorem 2.2.3 and Remark 2.2.4, and v_2 is provided by Corollary 2.2.12. The only requirement for v_3 is that G admits an exponential moment of order v_3 (see Remark 2.2.9), and v_4 comes from Lemma 2.2.15. ♢

The rest of this paper is organized as follows: in Section 2.2, we presents some heuristics of our method, and we provide tools to get lower bounds for the convergence speed to equilibrium of the PDMP, considering three successive phases (the age coalescence in Section 2.2.2, the Wasserstein coupling in Section 2.2.3 and the total variation coupling in Section 2.2.4). Afterwards, we will use those bounds in Section 2.3.1 to prove Theorem 2.1.1. Finally, a particular and convenient case is treated in Section 2.3.2. Indeed, if the inter-intake times have an exponential distribution, better speeds of convergence may be provided.

2.2 Explicit speeds of convergence

In this section, we draw our inspiration from coupling methods provided in [CMP10, BCG+13b] (for the *Transmission Control Protocol* (TCP) window size process), and in [Lin86, Lin92] (for renewal processes). Two other standard references for coupling methods are [Res92, Asm03]. The sequel provides not only existence and uniqueness of an invariant probability measure for (P_t) (by consequence of our result, but it could also be proved by Foster-Lyapounov methods, which may require some slightly different assumptions, see [MT93a] or [Hai10] for example) but also explicit exponential speeds of convergence to equilibrium for the total variation distance. The task is similar for convergence in Wasserstein distances.

Let us now briefly recall the definitions of the distances we use (see [Vil09] for details). Let $\mu, \tilde{\mu}$ be two probability measures on $\mathbb{R}^d$ (we denote by $\mathscr{M}(E)$ the set of probability measures on E). Then, we call coupling of μ and $\tilde{\mu}$ any probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ whose marginals are μ and $\tilde{\mu}$, and we denote by $\Gamma(\mu, \tilde{\mu})$ the set of all the couplings of μ and $\tilde{\mu}$. Let $p \in [1, +\infty)$; if we denote by $\mathscr{L}(X)$ the law of any random vector X, the Wasserstein distance between μ and $\tilde{\mu}$ is defined by

$$W_p(\mu, \tilde{\mu}) = \inf_{\mathscr{L}(X, \tilde{X}) \in \Gamma(\mu, \tilde{\mu})} \mathbb{E}\left[\|X - \tilde{X}\|^p\right]^{\frac{1}{p}}. \tag{2.2.1}$$

Similarly, the total variation distance between $\mu, \tilde{\mu} \in \mathscr{M}(\mathbb{R}^d)$ is defined by

$$\|\mu - \tilde{\mu}\|_{TV} = \inf_{\mathscr{L}(X, \tilde{X}) \in \Gamma(\mu, \tilde{\mu})} \mathbb{P}(X \neq \tilde{X}). \tag{2.2.2}$$

Moreover, we note (for real-valued random variables) $\mu \overset{\mathscr{L}}{\leq} \tilde{\mu}$ if $\mu((-\infty,x]) \geq \tilde{\mu}((-\infty,x])$ for all $x \in \mathbb{R}$. By a slight abuse of notation, we may use the previous notations for random variables instead of their distributions. It is known that both convergence in W_p and in total variation distance imply convergence in distribution. Observe that any arbitrary coupling provides an upper bound for the left-hand side terms in (2.2.1) and (2.2.2). The classical egality below is easy to show, and will be used later to provide a useful coupling; assuming that μ and $\tilde{\mu}$ admit f and $\tilde{f}$ for respective densities, there exists a coupling $\mathscr{L}(X,\tilde{X}) \in \Gamma(\mu,\tilde{\mu})$ such that

$$\mathbb{P}(X = \tilde{X}) = \int_{\mathbb{R}} f(x) \wedge \tilde{f}(x)dx. \tag{2.2.3}$$

Thus,

$$\|\mu - \tilde{\mu}\|_{TV} = 1 - \int_{\mathbb{R}} f(x) \wedge \tilde{f}(x)dx = \frac{1}{2}\int_{\mathbb{R}} |f(x) - \tilde{f}(x)|dx. \tag{2.2.4}$$

2.2.1 Heuristics

If, given a coupling $(Y,\tilde{Y}) = \big((X,\Theta,A),(\tilde{X},\tilde{\Theta},\tilde{A})\big)$, we can explicitly control the distance of their distributions at time t regarding their distance at time 0, and if $\mathscr{L}(\tilde{Y}_0)$ is the invariant probability measure, then we control the distance between $\mathscr{L}(Y_t)$ and this distribution. Formally, let $Y = (X,\Theta,A)$ and $\tilde{Y} = (\tilde{X},\tilde{\Theta},\tilde{A})$ be two PDMPs generated by (2.1.1) such as $Y_0 \overset{\mathscr{L}}{=} \mu_0$ and $\tilde{Y}_0 \overset{\mathscr{L}}{=} \tilde{\mu}_0$. Denote by μ (resp. $\tilde{\mu}$) the law of Y (resp. $\tilde{Y}$). We call coalescing time of Y and $\tilde{Y}$ the random variable

$$\tau = \inf\{t \geq 0 : \forall s \geq 0, Y_{t+s} = \tilde{Y}_{t+s}\}.$$

Note that τ is not, a priori, a stopping time (w.r.t. the natural filtration of Y and $\tilde{Y}$). It is easy to check from (2.2.2) that, for $t > 0$,

$$\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq \mathbb{P}(Y_t \neq \tilde{Y}_t) \leq \mathbb{P}(\tau > t). \tag{2.2.5}$$

As a consequence, the main idea is to fix $t > 0$ and to exhibit a coupling $(Y,\tilde{Y})$ such that $\mathbb{P}(\tau \geq t)$ is exponentially decreasing. Let us now present the coupling we shall use to that purpose. The justifications will be given in Sections 2.2.2, 2.2.3 and 2.2.4.

- Phase 1: Ages coalescence (from 0 to t_1)
  If X and $\tilde{X}$ jump separately, it is difficult to control their distance, because we can not control the height of their jumps (if F is not trivial). The aim of the first phase is to force the two processes to jump at the same time once; then, it is possible to choose a coupling with exactly the same jump mechanisms, which makes that the first jump is the coalescing time for A and $\tilde{A}$. Moreover, the randomness of U does not affect the strategy anymore afterwards, since it can be the same for both processes. Similarly, the randomness of Θ does not matter anymore. Finally, note that, if ζ is constant, it is always possible to make the processes jump at the same time, and the length of this phase exactly follows an exponential law of parameter $\zeta(0)$.

- Phase 2: Wasserstein coupling (from t_1 to t_2)
  Once there is coalescence of the ages, it is time to bring X and $\tilde{X}$ close to each other. Since we can give the same metabolic parameter and the same jumps at the same time for each process, knowing the distance and the metabolic parameter after the intake, the distance is deterministic until the next jump. Consequently, the distance between X and $\tilde{X}$ at time $s \in [t_1, t_2]$ is
  $$|X_s - \tilde{X}_s| = |X_{t_1} - \tilde{X}_{t_1}| \exp\left(-\int_{t_1}^{s} \Theta_r dr\right).$$
- Phase 3: Total variation coupling (from t_2 to t)
  If X and $\tilde{X}$ are close enough at time t_2, which is the purpose of phase 2, we have to make them jump simultaneously - again - but now at the same point. This can be done since F has a density. In this case, we have $\tau \leq t$; if this is suitably done, then $\mathbb{P}(\tau \leq t)$ is close to 1 and the result is given by (2.2.5).

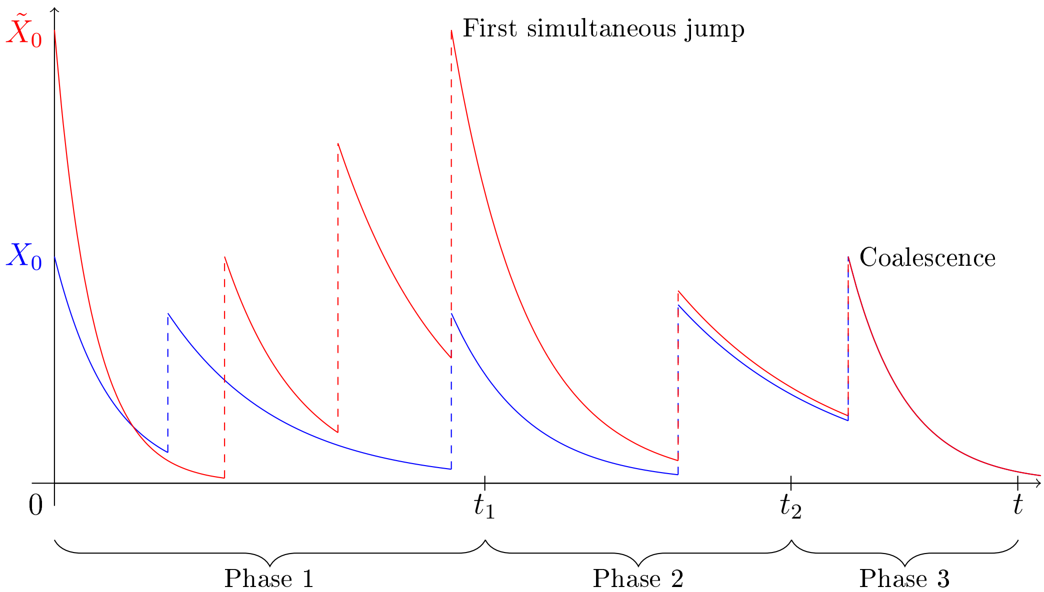


Figure 2.2.1 – Expected behaviour of the coupling.

This coupling gives us a good control of the total variation distance of Y and $\tilde{Y}$, and it can also provide an exponential convergence speed in Wasserstein distance if we set $t_2 = t$; this control is expressed with explicit rates of convergence in Theorem 2.1.1.

2.2.2 Ages coalescence

As explained in Section 2.2.1, we try to bring the ages A and $\tilde{A}$ to coalescence. Observe that knowing the dynamics of $Y = (X, \Theta, A)$, A is a PDMP with infinitesimal generator

$$\mathcal{A}\varphi(a) = \partial_a \varphi(a) + \zeta(a)[\varphi(0) - \varphi(a)], \tag{2.2.6}$$

so, for now, we will focus only on the age processes A and $\tilde{A}$, which is a classical renewal process. The reader may refer to [Fel71] or [Asm03] for deeper insights about renewal theory. Since ΔT_1 does not follow a priori the distribution G, A is a delayed renewal process; anyway this does not affect the sequel, since our method requires to wait for the first jump to occur.

Let $\mu_0, \tilde{\mu}_0 \in \mathscr{M}(\mathbb{R}_+)$. Denote by $(A, \tilde{A})$ the Markov process generated by the following infinitesimal generator:

$$\mathcal{A}_2\varphi(a,\tilde{a}) = \partial_a\varphi(a,\tilde{a})+\partial_{\tilde{a}}\varphi(a,\tilde{a})+[\zeta(a)-\zeta(\tilde{a})][\varphi(0,\tilde{a})-\varphi(a,\tilde{a})]+\zeta(\tilde{a})[\varphi(0,0)-\varphi(a,\tilde{a})] \tag{2.2.7}$$

if $\zeta(a) \geq \zeta(\tilde{a})$, and with a symmetric expression if $\zeta(a) < \zeta(\tilde{a})$, and such as $A_0 \overset{\mathscr{L}}{=} \mu_0$ and $\tilde{A}_0 \overset{\mathscr{L}}{=} \tilde{\mu}_0$. If $\varphi(a, \tilde{a})$ does not depend on a or on $\tilde{a}$, one can easily check that (2.2.7) reduces to (2.2.6), which means that $(A, \tilde{A})$ is a coupling of μ and $\tilde{\mu}$. Moreover, it is easy to see that, if a common jump occurs for A and $\tilde{A}$, every following jump will be simultaneous (since the term $\zeta(a) - \zeta(\tilde{a})$ will stay equal to 0 in $\mathcal{A}_2$). Note that, if ζ is a constant function, then this term is still equal to 0 and the first jump is common. Last but not least, since ζ is non-decreasing, only two phenomenons can occur: the older process jumps, or both jump together (in particular, if the younger process jumps, the other one jumps as well).

Our goal in this section is to study the time of the first simultaneous jump which will be, as previously mentionned, the coalescing time of A and $\tilde{A}$; by definition, here, it is a stopping time. Let

$$\tau_A = \inf\{t \geq 0 : A_t = \tilde{A}_t\} = \inf\{t \geq 0 : \forall s \geq 0, A_{t+s} = \tilde{A}_{t+s}\}.$$

Let

$$\begin{cases} a & = \inf\{t \geq 0 : \zeta(t) > 0\} \in [0, +\infty), \\ d & = \sup\{t \geq 0 : \zeta(t) < +\infty\} \in (0, +\infty]. \end{cases}$$

Remark 2.2.1: Note that assumption (H3) guarantees that $\inf \zeta = \zeta(a)$ and $\sup \zeta = \zeta(d^-)$. Moreover, if $d < +\infty$, then $\zeta(d^-) = +\infty$ since G admits a density. Indeed, the following relation is a classical result:

$$\int_0^{\Delta T} \zeta(s)ds \overset{\mathscr{L}}{=} \mathscr{E}(1),$$

which is impossible if $d < +\infty$ and $\zeta(d^-) < +\infty$. A slight generalisation of our model would be to use truncated random variables of the form $\Delta T \wedge C$ for a deterministic constant C; then, their common distribution would not admit a density anymore, but the mechanisms of the process would be similar. In that case, it is possible that $d < +\infty$ and $\zeta(d^-) < +\infty$, but the rest of the method remains unchanged. ♢

First, let us give a good and simple stochastic bound for τ_A in a particular case.

Proposition 2.2.2

If $\zeta(0) > 0$ then the following stochastic inequality holds:

$$\tau_A \overset{\mathscr{L}}{\leq} \mathscr{E}(\zeta(0)).$$

Proof: It is possible to rewrite (2.2.7) as follows:

$$\begin{aligned}\mathcal{A}_2\varphi(a,\tilde{a}) = \partial_a\varphi(a,\tilde{a}) + \partial_{\tilde{a}}\varphi(a,\tilde{a}) &+ [\zeta(a)-\zeta(\tilde{a})][\varphi(0,\tilde{a})-\varphi(a,\tilde{a})]\\ &+ [\zeta(\tilde{a})-\zeta(0)][\varphi(0,0)-\varphi(a,\tilde{a})]\\ &+ \zeta(0)[\varphi(0,0)-\varphi(a,\tilde{a})],\end{aligned}$$

for $\zeta(a) \geq \zeta(\tilde{a})$. This decomposition of (2.2.7) indicates that three independent phenomenons can occur for A and $\tilde{A}$ with respective hazard rates $\zeta(a)-\zeta(\tilde{a}), \zeta(\tilde{a})-\zeta(0)$ and $\zeta(0)$. We have a common jump in the last two cases and, in particular, the inter-arrival times of the latter follow a distribution $\mathscr{E}(\zeta(0))$ since the rate is constant. Thus, we have $\tau_A \overset{\mathscr{L}}{\leq} \mathscr{E}(\zeta(0))$. □

To rephrase this result, the age coalescence occurs stochastically faster than an exponential law. This relies only on the fact that the jump rate is bounded from below, and it is trickier to control the speed of coalescence if ζ is allowed to be arbitrarily close to 0. This is the purpose of the following theorem.

Theorem 2.2.3

Assume that $\inf \zeta = 0$. *Let* $\varepsilon > \frac{a}{2}$. *Let* $b, c \in (a,d)$ *such that* $\zeta(b) > 0$ *and* $c > b+\varepsilon$.

i) If $\frac{3a}{2} < d < +\infty$, *then*

$$\tau_A \overset{\mathscr{L}}{\leq} c + (2H-1)\varepsilon + \sum_{i=1}^{H} (d-\varepsilon)G^{(i)},$$

where $H, G^{(i)}$ *are independent random variables of geometric law and* $G^{(i)}$ *are i.i.d.*

ii) If $d = +\infty$ *and* $\zeta(d^-) < +\infty$, *then*

$$\tau_A \overset{\mathscr{L}}{\leq} \sum_{i=1}^{H}\sum_{j=1}^{G^{(i)}} \left(b + E^{(i,j)}\right),$$

where $H, G^{(i)}, E^{(i,j)}$ *are independent random variables,* $G^{(i)}$ *are i.i.d. with geometric law,* $E^{(i,j)}$ *are i.i.d. with exponential law and* $\mathscr{L}(H)$ *is geometric.*

iii) If $d = +\infty$ *and* $\zeta(d^-) = +\infty$, *then*

$$\tau_A \overset{\mathscr{L}}{\leq} c - \varepsilon + \sum_{i=1}^{H}\left(2\varepsilon + \sum_{j=1}^{G^{(i)}}\left(c-\varepsilon+E^{(i,j)}\right)\right),$$

where $H, G^{(i)}, E^{(i,j)}$ *are independent random variables,* $G^{(i)}$ *are i.i.d. with geometric law,* $E^{(i,j)}$ *are i.i.d. with exponential law and* $\mathscr{L}(H)$ *is geometric.*

Furthermore, the parameters of the geometric and exponential laws are explicit in terms of the parameters ε, a, b, c *and* d *(see the proof for details).*

Remark 2.2.4: Such results may look technical, but above all they allow us to know that the distribution tail of τ_A is exponentially decreasing (just like the geometric or exponential laws). If G is known (or equivalently, ζ), Theorem 2.2.3 provides a quantitative exponential bound for the tail. For instance, in case *i)*, if $\mathscr{L}(G^{(i)}) = \mathscr{G}(p_1)$ and $\mathscr{L}(H) = \mathscr{G}(p_2)$, then τ_A admits exponential moments strictly less than $-\frac{1}{2}\min\left(\frac{\log(1-p_2)}{2\varepsilon}, \frac{\log(1-p_1p_2)}{d-\varepsilon}\right)$, since H and $\sum_{i=1}^{H} G^{(i)}$ are (non-independent) random variables with respective exponential moments $-\log(1-p_2)^-$ and $-\log(1-p_1p_2)^-$. $\diamondsuit$

Remark 2.2.5: In the case *i)*, we make the technical assumption that $d \geq \frac{3a}{2}$; this is not compulsory and the results are basically the same, but we cannot use our technique. It comes from the fact that it is really difficult to make the two processes jump together if $d - a$ is small. Without such an assumption, one may use the same arguments with a greater number of jumps, in order to gain room for the jump time of the older process. Provided that the distribution G is spread-out, it is possible to bring the coupling to coalescence (see Theorem VII.2.7 in [Asm03]) but it is more difficult to obtain quantitative bounds. $\diamondsuit$

Remark 2.2.6: Even if Theorem 2.2.3 holds for any set of parameters (recall that a and d are fixed), it can be optimized by varying ε, b and c, depending on ζ. One should choose ε to be small regarding the length of the jump domain $[b, c]$ (which should be large, but with a small variation of ζ to maximize the common jump rate). $\diamondsuit$

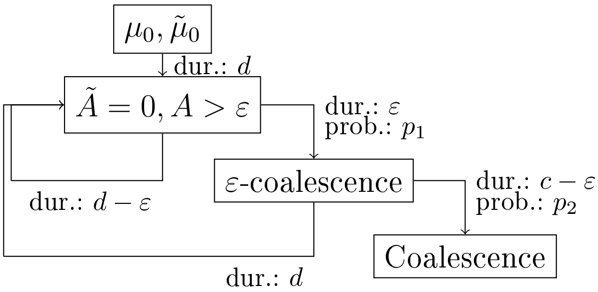
Proof of Theorem 2.2.3: First and foremost, let us prove *i)*. We recall that the processes A and $\tilde{A}$ jump necessarily to 0. The method we are going to use here will be applied to the other cases with a few differences. The idea is the following: try to make the distance between A and $\tilde{A}$ smaller than ε (which will be called a ε-coalescence), and then make the processes jump together where we can quantify their jump speed (i.e. in a domain where the jump rate is bounded, so that the simultaneous jump is stochastically bounded between two exponential laws). We make the age processes jump together in the domain $[b, c]$, whose length must be greater than ε; since $\varepsilon \geq a/2$ and $[b, c] \subset (a, d)$, this is possible only if $d > \frac{3a}{2}$. Then, we use the following algorithm:

- Step 1: Wait for a jump, so that one of the processes (say $\tilde{A}$) is equal to 0. The length of this step is less than $d < +\infty$ by definition of d.

- Step 2: If there is not yet ε-coalescence (say we are at time T), then $A_T > \varepsilon$. We want A to jump before a time ε, so that the next jump implies ε-coalescence. This probability is $1 - \exp\left(-\int_0^\varepsilon \zeta(A_T + s)ds\right)$, which is greater than the probability p_1 that a random variable following an exponential law of parameter $\zeta\left(\varepsilon + \frac{a}{2}\right)$ is less than $\varepsilon - \frac{a}{2}$. It corresponds to the probability of A jumping between $\frac{a+2\varepsilon}{2}$ and 2ε.

- Step 3: There is a ε-coalescence. Say $\tilde{A} = 0$ and $A \leq \varepsilon$. Recall that if the younger process jumps, the jump is common. So, if A does not jump before a time b, which probability is greater than $\exp\left(-b\zeta(b+\varepsilon)\right)$, and then $\tilde{A}$ jumps before a time $c - b - \varepsilon$, with a probability greater than $1 - \exp\left(-\left(c - b - \varepsilon\right)\zeta(b)\right)$, then coalescence occurs; else go back to Step 2.

The previous probabilities can be rephrased with the help of exponential laws:



Step 3 leads to coalescence with the help of the arguments mentionned before, using the expression (2.2.7) of $\mathcal{A}_2$. Simple computations show that

$$\begin{aligned}
p_1 &= 1 - \exp\left(-\left(\varepsilon - \frac{a}{2}\right)\zeta\left(\varepsilon + \frac{a}{2}\right)\right), \\
p_2 &= \exp\left(-b\zeta(b+\varepsilon)\right)\big(1 - \exp\left(-\left(c - b - \varepsilon\right)\zeta(b)\right)\big).
\end{aligned}$$

Let $G^{(i)} \overset{\mathscr{L}}{=} \mathscr{G}(p_1)$ be i.i.d. and $H \overset{\mathscr{L}}{=} \mathscr{G}(p_2)$ Then the following stochastic inequality holds:

$$\begin{aligned}
\tau_A &\overset{\mathscr{L}}{\leq} d + (d-\varepsilon)(G^{(1)} - 1) + \varepsilon + \mathbb{1}_{\{H \geq 2\}} \sum_{i=2}^{H} \big(d + (d-\varepsilon)(G^{(i)} - 1) + \varepsilon\big) + (c - \varepsilon) \\
&\overset{\mathscr{L}}{\leq} c + (2H-1)\varepsilon + \sum_{i=1}^{H} (d-\varepsilon) G^{(i)}.
\end{aligned}$$

Now, we prove *ii)*. We make the processes jump simultaneously in the domain $[b, +\infty)$ with the following algorithm:

- Step 1: Say A is greater than $\tilde{A}$. We want it to wait for $\tilde{A}$ to be in domain $[b, +\infty)$. In the worst scenario, it has to wait a time b, with a hazard rate less than $\zeta(d^-) < +\infty$. This step lasts less than a geometrical number of times b.

- Step 2: Once the two processes are in the jump domain, two phenomenons can occur: common jump with hazard rate greater than $\zeta(b)$ or jump of the older one with hazard rate less than $\zeta(d^-)$. The first jump occurs with a rate less than $\zeta(d^-)$ and is a simultaneous jump with probability greater than $\frac{\zeta(b)}{\zeta(d^-)}$. If there is no common jump, go back to Step 1.

Let

$$p_1 = \mathrm{e}^{-b\zeta(d^-)}, \quad p_2 = \frac{\zeta(b)}{\zeta(d^-)}.$$

Let $G^{(i)} \overset{\mathscr{L}}{=} \mathscr{G}(p_1)$ be i.i.d.,$H \overset{\mathscr{L}}{=} \mathscr{G}(p_2)$ and $E^{(i,j)} \overset{\mathscr{L}}{=} \mathscr{E}(\zeta(b))$ be i.i.d. Then the following stochastic inequality holds:

$$\begin{aligned}
\tau_A &\overset{\mathscr{L}}{\leq} \sum_{j=2}^{G^{(1)}} \left(b + E^{(1,j)}\right) + b + \mathbb{1}_{\{H \geq 2\}} \sum_{i=2}^{H} \left(E^{(i,1)} + \sum_{j=2}^{G^{(i)}} \left(b + E^{(i,j)}\right) + b \right) + E^{(1,1)} \\
&\overset{\mathscr{L}}{\leq} \sum_{i=1}^{H} \sum_{j=1}^{G^{(i)}} \left(b + E^{(i,j)}\right).
\end{aligned}$$

Let us now prove *iii)*. We do not write every detail here, since this case is a combination of the two previous cases (wait for a ε-coalescence, then bring the processes to coalescence using stochastic inequalities involving exponential laws). Let

$$\begin{aligned}
p_1 &= 1 - \exp\left(-\left(\varepsilon - \frac{a}{2}\right)\zeta\left(\varepsilon + \frac{a}{2}\right)\right), \\
p_2 &= \frac{\zeta(b)}{\zeta(c)} \exp\left(-b\zeta(b+\varepsilon)\right)\left(1 - \exp\left(-(c-b-\varepsilon)\zeta(b)\right)\right).
\end{aligned}$$

Let $G^{(i)} \overset{\mathscr{L}}{=} \mathscr{G}(p_1)$ be i.i.d., $H \overset{\mathscr{L}}{=} \mathscr{G}(p_2)$ and $E^{(i,j)} \overset{\mathscr{L}}{=} \mathscr{E}(\zeta(c))$ be i.i.d. Then the following stochastic inequality holds

$$\begin{aligned}
\tau_A &\overset{\mathscr{L}}{\leq} c + E^{(1,1)} + \varepsilon + \sum_{j=2}^{G^{(1)}} \left(c - \varepsilon + E^{(1,j)}\right) + (c - \varepsilon) \\
&\quad + \sum_{i=2}^{H} \left(c + E^{(i,1)} + \varepsilon + \sum_{j=2}^{G^{(i)}} \left(c - \varepsilon + E^{(i,j)}\right) \right) \\
&\overset{\mathscr{L}}{\leq} c - \varepsilon + \sum_{i=1}^{H} \left(2\varepsilon + \sum_{j=1}^{G^{(i)}} \left(c - \varepsilon + E^{(i,j)}\right) \right).
\end{aligned}$$

□

2.2.3 Wasserstein coupling

Let $\mu_0, \tilde{\mu}_0 \in \mathscr{M}(\mathbb{R}_+)$. Denote by $(Y, \tilde{Y}) = (X, \Theta, A, \tilde{X}, \tilde{\Theta}, \tilde{A})$ the Markov process generated by the following infinitesimal generator:

$$\begin{aligned}
\mathcal{L}_2\varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a}) &= \int_{u=0}^{\infty} \int_{\theta'=0}^{\infty} \Big([\zeta(a) - \zeta(\tilde{a})]\left[\varphi(x+u, \theta', 0, \tilde{x}, \tilde{\theta}, \tilde{a}) - \varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a})\right] \\
&\quad + \zeta(\tilde{a})\left[\varphi(x+u, \theta', 0, \tilde{x}+u, \theta', 0) - \varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a})\right]\Big) H(d\theta')F(du) \\
&\quad - \theta x \partial_x \varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a}) - \tilde{\theta}\tilde{x}\partial_x \varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a}) \\
&\quad + \partial_a \varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a}) + \partial_{\tilde{a}} \varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a})
\end{aligned} \tag{2.2.8}$$

if $\zeta(a) \geq \zeta(\tilde{a})$, and with a symmetric expression if $\zeta(a) < \zeta(\tilde{a})$, and with $Y_0 \overset{\mathscr{L}}{=} \mu_0$ and $\tilde{Y}_0 \overset{\mathscr{L}}{=} \tilde{\mu}_0$. As in the previous section, one can easily check that Y and $\tilde{Y}$ are generated by (2.1.1) (so $(Y, \tilde{Y})$ is a coupling of μ and $\tilde{\mu}$). Moreover, if we choose $\varphi(x, \theta, a, \tilde{x}, \tilde{\theta}, \tilde{a}) = \psi(a, \tilde{a})$ then (2.2.8) reduces to (2.2.7), which means that the results of the previous section still hold for the age processes embedded in a coupling generated by (2.2.8). As explained in Section 2.2.2, if Y and $\tilde{Y}$ jump simultaneously, then they will always jump together afterwards. After the age coalescence, the metabolic parameters and the contaminant quantities are the same for Y and $\tilde{Y}$. Thus, it is easy to deduce the following lemma, whose proof is straightforward with the previous arguments.

Lemma 2.2.7

Let $(Y, \tilde{Y})$ be generated by $\mathcal{L}_2$ in (2.2.8). If $A_{t_1} = \tilde{A}_{t_1}$ and $\Theta_{t_1} = \tilde{\Theta}_{t_1}$, then, for $t \geq t_1$,

$$A_t = \tilde{A}_t, \quad \Theta_t = \tilde{\Theta}_t.$$

Moreover,

$$|X_t - \tilde{X}_t| = |X_{t_1} - \tilde{X}_{t_1}| \exp\left(-\int_{t_1}^t \Theta_s ds\right).$$

From now on, let $(Y, \tilde{Y})$ be generated by $\mathcal{L}_2$ in (2.2.8). We need to control the Wasserstein distance of X_t and $\tilde{X}_t$; this is done in the following theorem. The reader may refer to [Asm03] for a definition of the direct Riemann-integrability (d.R.i.); one may think at first of "non-negative, integrable and asymptotically decreasing". In the sequel, we denote by ψ_J the Laplace transform of any positive measure J: $\psi_J(u) = \int_{\mathbb{R}} \mathrm{e}^{ux} J(dx)$.

Theorem 2.2.8

Let $p \geq 1$. Assume that $A_0 = \tilde{A}_0$ and $\Theta_0 = \tilde{\Theta}_0$.

i) If $G = \mathscr{E}(\lambda)$ (i.e. ζ is constant, equal to λ) then,

$$\mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\right] \leq \exp\left(-\lambda(1 - \mathbb{E}\left[\mathrm{e}^{-p\Theta_1 T_1}\right])t\right). \tag{2.2.9}$$

ii) Let

$$J(dx) = \mathbb{E}\left[\mathrm{e}^{-p\Theta_1 x}\right] G(dx), \quad w = \sup\{u \in \mathbb{R} : \psi_J(u) < 1\}.$$

If $\sup\{u \in \mathbb{R} : \psi_J(u) < 1\} = +\infty$, let w be any positive number. Then for all $\varepsilon > 0$, there exists $C > 0$ such that

$$\mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\right] \leq Ce^{-(w-\varepsilon)t}. \tag{2.2.10}$$

Furthermore, if $\psi_J(w) < 1$ and $\psi_G(w) < +\infty$, or if $\psi_J(w) \leq 1$ and the function $t \mapsto \mathrm{e}^{wt}\mathbb{E}\left[\mathrm{e}^{-p\Theta_1 t}\right] G((t, +\infty))$ is directly Riemann-integrable, then there exists $C > 0$ such that

$$\mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\right] \leq Ce^{-wt}. \tag{2.2.11}$$

Remark 2.2.9: Note that $w > 0$ by (H3), since the probability measure G admits an exponential moment. Indeed, there exist $l, m > 0$ such that, for $t \geq l, \zeta(t) \geq m$. Hence $G \overset{\mathscr{L}}{\leq} l + \mathscr{E}(m)$, and $\psi_G(u) \leq \mathrm{e}^{ul} + m(m-u)^{-1} < +\infty$ for $u < m$. In particular, if $\sup \zeta = +\infty$, the domain of ψ_G is the whole real line, and (2.2.11) holds. $\diamondsuit$

Remark 2.2.10: The previous theorem provides a speed of convergence toward 0 for $\mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\right]$ when $t \to +\infty$ under various assumptions. To prove it, we turn to the renewal theory (for a good review, see [Asm03]), which has already been widely studied. Here, we link the boundaries we obtained to the parameters of our model. $\diamondsuit$

Remark 2.2.11: If $\sup\{u \in \mathbb{R} : \psi_J(u) < 1\} = +\infty$, Theorem 2.2.8 asserts that, for any $w > 0$, there exists $C > 0$ such that $Z \leq C\mathrm{e}^{-wt}$, which means its decay is faster than any exponential rate. Moreover, note that a sufficient condition for $t \mapsto \mathrm{e}^{wt}\mathbb{E}\left[\mathrm{e}^{-p\Theta t}\right]\mathbb{P}(\Delta T > t)$ to be d.R.i. is that there exists $\varepsilon > 0$ such that $\psi_G(w+\varepsilon) < +\infty$. Indeed,

$$\mathrm{e}^{wt}\mathbb{E}[\mathrm{e}^{-p\Theta t}]\mathbb{P}(\Delta T > t) \leq \mathrm{e}^{wt}\mathbb{E}[\mathrm{e}^{-p\Theta t}]\mathrm{e}^{-(w+\varepsilon)t}\psi_G(w+\varepsilon) \leq \psi_G(w+\varepsilon)\mathrm{e}^{-\varepsilon t},$$

and the right-hand side is d.R.i. $\diamondsuit$

Proof of Theorem 2.2.8: In this context, $\mathscr{L}(\Delta T_1) \overset{\mathscr{L}}{\leq} G$; it is harmless to assume that $\mathscr{L}(\Delta T_1) \overset{\mathscr{L}}{=} G$, since this assumptions only slows the convergence down. Then, denote by Θ and ΔT two random variables distributed according to H and G respectively. Let us prove *i)*; in this particular case, since ζ is constant equal to λ, $N_t \overset{\mathscr{L}}{=} \mathscr{P}(\lambda t)$, so

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\right] &= \mathbb{E}\left[\exp\left(-\mathbb{1}_{\{N_t\geq 1\}}\sum_{i=1}^{N_t} p\Theta_i\Delta T_i - p\Theta_{N_t+1}(t-T_{N_t})\right)\right] \\
&\leq \mathbb{E}\left[\exp\left(-\mathbb{1}_{\{N_t\geq 1\}}\sum_{i=1}^{N_t} p\Theta_i\Delta T_i\right)\right] \\
&\leq \mathbb{P}(N_t = 0) + \sum_{n=1}^{\infty}\mathbb{E}\left[\exp\left(-\sum_{i=1}^{n} p\Theta_i\Delta T_i\right)\right]\mathbb{P}(N_t = n) \\
&\leq \mathrm{e}^{-\lambda t} + \sum_{n=1}^{\infty}\mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathbb{E}\left[\mathrm{e}^{-p\Theta\Delta T}\right]^n \\
&\leq \exp\left(-\lambda(1-\mathbb{E}[\mathrm{e}^{-p\Theta\Delta T}])t\right).
\end{aligned}$$

Now, let us prove *ii)*. Let $Z(t) = \mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\right]$; we have

$$\begin{aligned}
Z(t) &= \mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\mathbb{1}_{\{T_1>t\}}\right] + \mathbb{E}\left[\exp\left(-\int_0^t p\Theta_s ds\right)\mathbb{1}_{\{T_1\le t\}}\right] \\
&= \mathbb{E}[\mathrm{e}^{-p\Theta t}]\mathbb{P}(\Delta T > t) + \int_0^t \mathbb{E}\left[\mathrm{e}^{-p\Theta x}\exp\left(-\int_x^t p\Theta_s ds\right)\right] G(dx) \\
&= \mathbb{E}[\mathrm{e}^{-p\Theta t}]\mathbb{P}(\Delta T > t) + \int_0^t \mathbb{E}\left[\mathrm{e}^{-p\Theta x}\right]\mathbb{E}\left[\exp\left(-\int_0^{t-x} p\Theta_s ds\right)\right] G(dx) \\
&= z(t) + J * Z(t),
\end{aligned}$$

where $z(t) = \mathbb{E}[\mathrm{e}^{-p\Theta t}]\mathbb{P}(\Delta T > t)$ and $J(dt) = \mathbb{E}[\mathrm{e}^{-p\Theta t}]G(dt)$. Since $J(\mathbb{R}) < 1$, the function Z satisfies the defective renewal equation

$$Z = z + J * Z.$$

Let $\varepsilon > 0$; the function ψ_J is well defined, continuous, non-decreasing on $(-\infty, w)$, and $\psi_J(w-\varepsilon) < 1$. Let

$$Z'(t) = \mathrm{e}^{(w-\varepsilon)t}Z(t), \quad z'(t) = \mathrm{e}^{(w-\varepsilon)t}z(t), \quad J'(dt) = \mathrm{e}^{(w-\varepsilon)t}J(dt).$$

It is easy to check that $J' * Z'(t) = \mathrm{e}^{(w-\varepsilon)t}J * Z(t)$, thus Z' satisfies the renewal equation

$$Z' = z' + J' * Z', \tag{2.2.12}$$

which is defective since $J'(\mathbb{R}) = \psi_{J'}(0) = \psi_J(w-\varepsilon) < 1$. Let

$$v = \sup\{u > 0 : \psi_G(u) < +\infty\}.$$

Since G admits exponential moments, $v \in (0, +\infty]$. If $w < v$,

$$\begin{aligned}
z'(t) &= \mathrm{e}^{(w-\varepsilon)t}\mathbb{E}\left[\mathrm{e}^{-p\Theta t}\right]\mathbb{P}\left(\mathrm{e}^{w\Delta T} > \mathrm{e}^{wt}\right) \le \mathrm{e}^{(w-\varepsilon)t}\mathbb{E}\left[\mathrm{e}^{-p\Theta t}\right]\psi_G(w)\mathrm{e}^{-wt} \\
&\le \psi_G(w)\mathrm{e}^{-\varepsilon t}\mathbb{E}\left[\mathrm{e}^{-p\Theta t}\right],
\end{aligned} \tag{2.2.13}$$

then $\lim_{t\to+\infty} z'(t) = 0$. If $v \le w$, temporarily set $\varphi(t) = \mathbb{E}\left[\exp\left((w - 2\varepsilon/3 - p\Theta - v)t\right)\right]$. Assume that $\mathbb{P}(w - 2\varepsilon/3 - p\Theta - v \ge 0) \ne 0$. Thus, if $\mathbb{P}(w - 2\varepsilon/3 - p\Theta - v > 0) > 0$, then $\lim_{t\to+\infty}\varphi(t) = +\infty$; else, $\lim_{t\to+\infty}\varphi(t) = \mathbb{P}(w - 2\varepsilon/3 - p\Theta - v = 0) > 0$. Anyway, there exist $t_0, M > 0$ such that for all $t \ge t_0, \varphi(t) \ge M$. It implies

$$\int_0^\infty \varphi(t)\mathrm{e}^{(v+\varepsilon/3)t}g(t)dt \ge M\int_{t_0}^\infty \mathrm{e}^{(v+\varepsilon/3)t}g(t)dt = +\infty,$$

since $\psi_G(v + \varepsilon/3) = +\infty$, which contradicts the fact that

$$\psi_J(w - \varepsilon/3) = \int_0^\infty \mathbb{E}\left[\exp\left((w - 2\varepsilon/3 - p\Theta - v)t\right)\right]\mathrm{e}^{(v+\varepsilon/3)t}g(t)dt < +\infty.$$

Thus, $\mathbb{P}(w - 2\varepsilon/3 - p\Theta - v < 0) = 1$ and $\lim_{t\to+\infty}\varphi(t) = 0$. Using the Markov inequality like for (2.2.13), we have

$$z'(t) \le \psi_G(v - \varepsilon/3)\mathbb{E}\left[\exp\left((w - 2\varepsilon/3 - p\Theta - v)t\right)\right] = \psi_G(v - \varepsilon/3)\varphi(t),$$

from which we deduce $\lim_{t\to+\infty} z'(t) = 0$. Using Proposition V.7.4 in [Asm03], Z' is bounded, so there exists $C > 0$ such that (2.2.10) holds. From [Asm03], note that the function Z' can be explicitly written as $Z' = \left(\sum_{n=0}^{\infty} (J')^{*n}\right) * z'$. Using this expression, it is possible to make C explicit, or at least to approximate it with numerical methods.

Eventually, we look at (2.2.12) in the case $\varepsilon = 0$. First, if $\psi_J(w) < 1$ and $\psi_G(w) < +\infty$, it is straightforward to apply the previous argument (since (2.2.12) remains defective and (2.2.13) still holds). Next, if $\psi_J(w) \leq 1$ and $z' : t \mapsto \mathrm{e}^{wt} z(t)$ is d.R.i., we can apply Theorem V.4.7 - the Key Renewal Theorem - or Proposition V.7.4 in [Asm03], whether $\psi_J(w) = 1$ or $\psi_J(w) < 1$. As a consequence, Z' is still bounded, and there still exists $C > 0$ such that (2.2.11) holds. □

The following corollary is of particular importance because it allows us to control the Wasserstein distance of the processes X and $\tilde{X}$ defined in (2.2.1).

Corollary 2.2.12

Let $p \geq 1$. Assume that $A_{t_1} = \tilde{A}_{t_1}, \Theta_{t_1} = \tilde{\Theta}_{t_1}$.

i) There exist $v > 0, C > 0$ such that, for $t \geq t_1$,

$$W_p(X_t, \tilde{X}_t) \leq C \exp\left(-v(t - t_1)\right) W_p(X_{t_1}, \tilde{X}_{t_1}).$$

ii) Furthermore, if ζ is a constant equal to λ then, for $t \geq t_1$,

$$W_p(X_t, \tilde{X}_t) \leq \exp\left(-\frac{\lambda}{p}(1 - \mathbb{E}[\mathrm{e}^{-p\Theta_1 T_1}])(t - t_1)\right) W_p(X_{t_1}, \tilde{X}_{t_1}).$$

Proof: By Markov property, assume w.l.o.g. that $t_1 = 0$. Under the notations of Theorem 2.2.8, note $v = p^{-1}(w - \varepsilon)$ for $\varepsilon > 0$, or even $v = p^{-1} w$ if $\psi_J(w) < 1$ and $\psi_G(w) < +\infty$, or $t \mapsto \mathrm{e}^{wt} \mathbb{E}\left[\mathrm{e}^{-p\Theta t}\right] \mathbb{P}(\Delta T > t)$ is directly Riemann-integrable. Thus, *i)* follows straightforwardly from (2.2.10) or (2.2.11) using Lemma 2.2.7. Relation *ii)* is obtained similarly from (2.2.9). □

2.2.4 Total variation coupling

Quantitative bounds for the coalescence of X and $\tilde{X}$, when A and $\tilde{A}$ are equal and X and $\tilde{X}$ are close, are provided in this section. We are going to use assumption (H1), which is crucial for our coupling method. Recall that we denote by f the density of F, which is the distribution of the jumps $U_n = X_{T_n} - X_{T_n^-}$. From (2.2.4), it is useful to set, for small ε,

$$\eta(\varepsilon) = 1 - \int_{\mathbb{R}} f(x) \wedge f(x - \varepsilon) dx = \frac{1}{2} \int_{\mathbb{R}} |f(x) - f(x - \varepsilon)| \, dx. \tag{2.2.14}$$

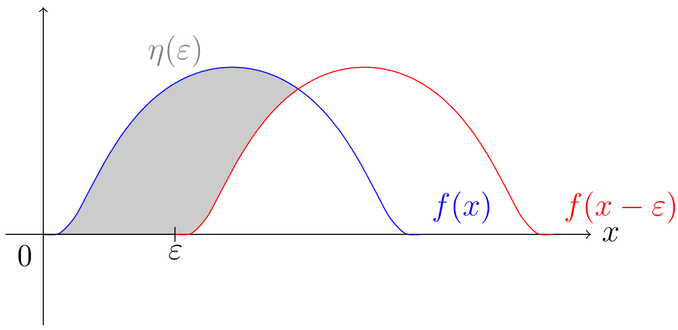


Figure 2.2.2 – Typical graph of η.

Definition 2.2.13

Assume that $A_t = \tilde{A}_t$. We call "TV coupling" the following coupling:

- *From t, let $(Y, \tilde{Y})$ be generated by $\mathcal{L}_2$ in (2.2.8) and make Y and $\tilde{Y}$ jump at the same time (say T).*
- *Then, knowing $(Y_{T^-}, \tilde{Y}_{T^-})$, use the coupling provided by (2.2.3) for $X_{T_-} + U$ and $\tilde{X}_{T^-} + \tilde{U}$.*

With the previous notations, conditioning on $\{X_{T^-}, \tilde{X}_{T^-}\}$, it is straightforward that $\mathbb{P}(X_T = \tilde{X}_T) \geq 1 - \eta\left(\left|X_{T^-} - \tilde{X}_{T^-}\right|\right)$. Let

$$\tau = \inf\{u \geq 0 : \forall s \geq u, Y_s = \tilde{Y}_s\}$$

be the coalescing time of Y and $\tilde{Y}$; from (2.2.4) and (2.2.14), one can easily check the following proposition.

Proposition 2.2.14

Let $\varepsilon > 0$. Assume that $A_{t_2} = \tilde{A}_{t_2}$, $\Theta_{t_2} = \tilde{\Theta}_{t_2}$ and $|X_{t_2} - \tilde{X}_{t_2}| \leq \varepsilon$. If $(Y, \tilde{Y})$ follows the TV coupling, then

$$\mathbb{P}\left(X_{T_{N_{t_2}+1}} \neq \tilde{X}_{T_{N_{t_2}+1}}\right) \leq \sup_{x\in[0,\varepsilon]} \eta(x).$$

This proposition is very important, since it enables us to quantify the probability to bring X and $\tilde{X}$ to coalescence (for small ε), and then (X, Θ, A) and $(\tilde{X}, \tilde{\Theta}, \tilde{A})$. With good assumptions on the density f (typically (H4a) or (H4b)), one can also easily control the term $\sup_{x\in[0,\varepsilon]} \eta(x)$; this is the point of the lemma below.

Lemma 2.2.15

Let $0 < \varepsilon < 1$. There exist $C, v > 0$ such that

$$\sup_{x\in[0,\varepsilon]} \eta(x) \leq C\varepsilon^v. \tag{2.2.15}$$

Proof: Assumptions (H4a) and (H4b) are crucial here. If (H4a) is fullfiled, which means η is Hölder, (2.2.15) is straightforward (and v is its Hölder exponent, since $\eta(0) = 0$). Otherwise, assume that (H4b) is true: f is h-Hölder, that is to say there exist $K, h > 0$ such that $|f(x) - f(y)| < K|x - y|^h$, and $\lim_{x\to+\infty} x^p f(x) = 0$ for some $p > 2$. Then, denote by D_ε the $(1 - \varepsilon^h)$-quantile of F, so that

$$\int_{D_\varepsilon}^{\infty} f(u)du = \varepsilon^h.$$

Then, we have, for all $x \leq \varepsilon$,

$$\begin{aligned} \eta(x) &= \frac{1}{2}\left(\int_0^{D_\varepsilon+1} |f(u) - f(u-x)|du + \int_{D_\varepsilon+1}^{\infty} |f(u) - f(u-x)|du\right) \\ &\leq \frac{1}{2}\int_0^{D_\varepsilon+1} |f(u) - f(u-x)|du + \frac{1}{2}\int_{D_\varepsilon+1}^{\infty} (f(u) + f(u-x))du \\ &\leq \left(K\frac{D_\varepsilon + 1}{2} + 1\right)\varepsilon^h. \end{aligned} \tag{2.2.16}$$

Now, let us control D_ε; there exists $C' > 0$ such that $f(x) \leq C'x^{-p}$. Then,

$$\int_{\left(\frac{C'}{(p-1)\varepsilon^h}\right)^{\frac{1}{p-1}}}^{\infty} f(x)dx \leq \int_{\left(\frac{(p-1)\varepsilon^h}{C'}\right)^{\frac{-1}{p-1}}}^{\infty} C'x^{-p}dx = \varepsilon^h,$$

so

$$D_\varepsilon \leq \left(\frac{C'}{(p-1)\varepsilon^h}\right)^{\frac{1}{p-1}}. \tag{2.2.17}$$

Denoting by

$$C = K\frac{\left(\frac{C'}{p-1}\right)^{\frac{1}{p-1}} + 1}{2} + 1, \quad v = h - \frac{h}{p-1},$$

the parameter v is positive because $p > 2$, and (2.2.15) follows from (2.2.16) and (2.2.17). □

2.3 Main results

In this section, we use the tools provided in Section 2.2 to bound the coalescence time of the processes and prove the main result of this paper, Theorem 2.1.1; some better results are also derived in a specific case. Two methods will be presented. The first one is general and may be applied in every case, whereas the second one uses properties of homogeneous Poisson processes, which is relevant only in the particular case where the inter-intake times follow an exponential distribution, and, a priori, cannot be used in other cases. From now on, let Y and $\tilde{Y}$ be two PDMPs generated by $\mathcal{L}$ in (2.1.1), with $\mathscr{L}(Y_0) = \mu_0$ and $\mathscr{L}(\tilde{Y}_0) = \tilde{\mu}_0$. Let t be a fixed positive real number, and, using (2.2.5), we aim at bounding $\mathbb{P}(\tau > t)$ from above ; recall that τ_A and τ are the respective coalescing times of the PDMPs A and $\tilde{A}$, and Y and $\tilde{Y}$. The heuristic is the following: the interval $[0, t]$ is splitted into three domains, where we apply the three results of Section 2.2.

- First domain: apply the strategy of Section 2.2.2 to get age coalescence.
- Second domain: move X and $\tilde{X}$ closer with $\mathcal{L}_2$, as defined in Section 2.2.3.
- Third domain: make X and $\tilde{X}$ jump at the same point, using the density of F and the TV coupling of Section 2.2.4.

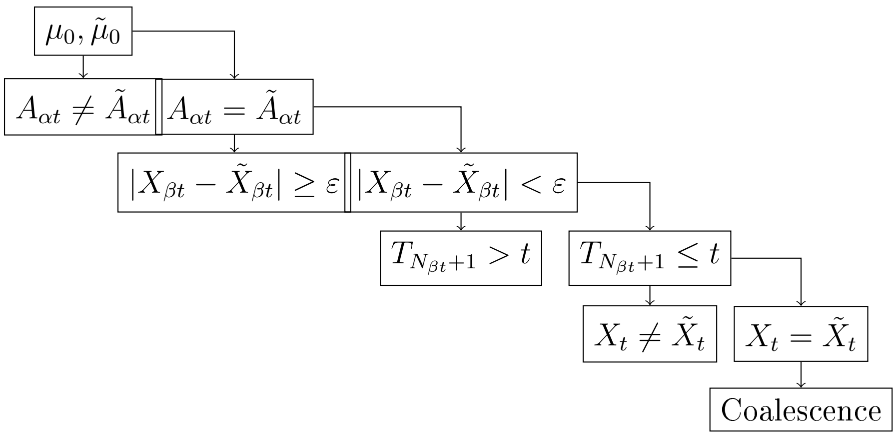
2.3.1 A deterministic division

The coupling method we present here bounds from above the total variation distance of the processes. The division of the interval $[0, t]$ will be deterministic, whereas it will be random in Section 2.3.2. To this end, let $0 < \alpha < \beta < 1$. The three domains will be $[0, \alpha t]$, $(\alpha t, \beta t]$ and $(\beta t, t]$. Now, we are able to prove Theorem 2.1.1. Recall that

$$\tau = \inf\{t \geq 0 : \forall s \geq 0, Y_{t+s} = \tilde{Y}_{t+s}\}$$

is the coalescing time of Y and $\tilde{Y}$, and τ_A is the coalescing time of A and $\tilde{A}$.

Proof of Theorem 2.1.1.*i)*: Let $\varepsilon > 0$. Let $(Y, \tilde{Y})$ be the coupling generated by $\mathcal{L}_2$ in (2.2.8) on $[0, \beta t]$ and the TV coupling on $(\beta t, t]$. Let us compute the probabilities of the following tree:



Recall from (2.2.5) that $\|\mu_0 P_t - \mu_0 P_t\|_{TV} \leq \mathbb{P}(\tau > t)$. Thus,

$$\begin{aligned}\mathbb{P}(\tau \leq t) \geq \mathbb{P}(\tau_A \leq \alpha t)\, \mathbb{P}\left(|X_{\beta t} - \tilde{X}_{\beta t}| < \varepsilon \middle| \tau_A \leq \alpha t\right) \\ \times \mathbb{P}\left(T_{N_{\beta t}+1} \leq t \middle| \tau_A \leq \alpha t, |X_{\beta t} - \tilde{X}_{\beta t}| < \varepsilon\right) \\ \times \mathbb{P}\left(\tau \leq t \middle| \tau_A \leq \alpha t, |X_{\beta t} - \tilde{X}_{\beta t}| < \varepsilon, T_{N_{\beta t}+1} \leq t\right). \qquad (2.3.1)\end{aligned}$$

First, by Theorem 2.2.3, we know that the distribution tail of τ_A is exponentially decreasing, since τ_A is a linear combination of random variables with exponential tails. Therefore,

$$\mathbb{P}(\tau_A > \alpha t) \leq C_1 \mathrm{e}^{-v_1 \alpha t},$$

where the parameters C_1 and v_1 are directly provided by Theorem 2.2.3 (see Remark 2.2.4). Now, conditioning on $\{\tau_A \le t\}$, using Corollary 2.2.12, there exist $C_2', v_2' > 0$ such that

$$\mathbb{P}\left(|X_{\beta t} - \tilde{X}_{\beta t}| \ge \varepsilon \middle| \tau_A \le \alpha t\right) \le \frac{W_1(X_{\beta t}, \tilde{X}_{\beta t})}{\varepsilon} \le \frac{W_1(X_{\alpha t}, \tilde{X}_{\alpha t})}{\varepsilon} C_2' \mathrm{e}^{-v_2'(\beta-\alpha)t}.$$

Let $U, \Delta T, \Theta$ be independent random variables of respective laws F, G, H, and say that any sum between i and j is equal to zero if $i > j$. We have

$$\begin{aligned}
\mathbb{E}\left[X_{\alpha t}\right] \le \mathbb{E}\left[X_{T_{N_{\alpha t}}}\right] &\le \mathbb{E}\left[X_0 \exp\left(-\sum_{k=2}^{N_{\alpha t}} \Theta_k \Delta T_k\right) + \sum_{i=1}^{N_{\alpha t}} U_i \exp\left(-\sum_{k=i+1}^{N_{\alpha t}} \Theta_k \Delta T_k\right)\right] \\
&\le \mathbb{P}(N_{\alpha t} = 0)\mathbb{E}[X_0] + \sum_{n=1}^{\infty} \mathbb{P}(N_{\alpha t} = n)\left(\mathbb{E}[X_0]\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]^{n-1} + \mathbb{E}[U]\sum_{k=0}^{n-1}\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]^k\right) \\
&\le \mathbb{E}[X_0] + \sum_{n=0}^{\infty} \mathbb{P}(N_{\alpha t} = n)\left(\frac{\mathbb{E}[X_0]\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]^n}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]} + \mathbb{E}[U]\frac{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]^n}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) \\
&\le \mathbb{E}[X_0] + \sum_{n=0}^{\infty} \mathbb{P}(N_{\alpha t} = n)\left(\frac{\mathbb{E}[X_0]}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]} + \frac{\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) \\
&\le \mathbb{E}[X_0]\left(1 + \frac{1}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) + \frac{\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}.
\end{aligned}$$

Hence,

$$\begin{aligned}
W_1(X_{\alpha t}, \tilde{X}_{\alpha t}) \le \mathbb{E}\left[X_{\alpha t} \vee \tilde{X}_{\alpha t}\right] &\le \mathbb{E}\left[X_{\alpha t}\right] + \mathbb{E}\left[\tilde{X}_{\alpha t}\right] \\
&\le (\mathbb{E}[X_0 + \tilde{X}_0])\left(1 + \frac{1}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) + \frac{2\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}.
\end{aligned}$$

Note $C_2 = \left((\mathbb{E}[X_0 + \tilde{X}_0])\left(1 + \frac{1}{\mathbb{E}[\mathrm{e}^{-\Theta\Delta T}]}\right) + \frac{2\mathbb{E}[U]}{1-\mathbb{E}[\mathrm{e}^{-\Theta\Delta T}]}\right) C_2'$. Recall that G admits an exponenital moment (see Remark 2.2.9). We have, using the Markov property, for all v_3 such that $\psi_G(v_3) < +\infty$:

$$\mathbb{P}\left(T_{N_{\beta t}+1} > t \middle| \tau_A \le \alpha t, |X_{\beta t} - \tilde{X}_{\beta t}| < \varepsilon\right) \le \mathbb{P}\left(\Delta T > (1-\beta)t\right) \le \psi_G(v_3)\mathrm{e}^{-v_3(1-\beta)t}.$$

Note $C_3 = \psi_G(v_3)$. Using Proposition 2.2.14 and Lemma 2.2.15, we have

$$\mathbb{P}\left(\tau > t \middle| \tau_A \le \alpha t, |X_{\beta t} - \tilde{X}_{\beta t}| < \varepsilon, T_{N_{\beta t}+1} \le t\right) \le \sup_{x\in[0,\varepsilon]} \eta(x) \le C_4 \varepsilon^{v_4'}.$$

The last step is to choose a correct ε to have exponential convergence for both the terms $\varepsilon^{-1} C_2 \mathrm{e}^{-v_2'(\beta-\alpha)t}$ and $C_4\varepsilon^{v_4'}$. The natural choice is to fix $\varepsilon = \mathrm{e}^{-v'(\beta-\alpha)t}$, for any $v' < v_2'$. Then, denoting by

$$v_2 = v_2' - v', \quad v_4 = v_4' v',$$

and using the equalities above, it is straightforward that (2.3.1) reduces to (2.1.2). □

Remark 2.3.1: Theorem 2.1.1 is very important and, above all, states that the exponential rate of convergence in total variation of the PDMP is larger than $\min(\alpha v_1, (\beta-\alpha)v_2, (1-\beta)v_3, (\beta-\alpha)v_4)$. If we choose

$$v' = \frac{v_2'}{1+v_4'}$$

in the proof above, the parameters v_2 and v_4 are equal; then, in order to have the maximal rate of convergence, one has to optimize α and β depending on v_1, v_2, v_3. $\diamondsuit$

Proof of Theorem 2.1.1.*ii)*: Let $(Y, \tilde{Y})$ be the coupling generated by $\mathcal{L}_2$ in (2.2.8). Note that

$$W_1(Y_t, \tilde{Y}_t) \leq \mathbb{E}\left[\|(X_t, \Theta_t, A_t) - (\tilde{X}_t, \tilde{\Theta}_t, \tilde{A}_t)\|\right] = \mathbb{E}[|X_t - \tilde{X}_t|] + \mathbb{E}[|\Theta_t - \tilde{\Theta}_t|] + \mathbb{E}[|A_t - \tilde{A}_t|].$$

Recall that $\mathbb{E}\left[X_{\alpha t}\right] \leq \mathbb{E}[X_0]\left(1 + \frac{1}{\mathbb{E}[\mathrm{e}^{-\Theta\Delta T}]}\right) + \frac{\mathbb{E}[U]}{1-\mathbb{E}[\mathrm{e}^{-\Theta\Delta T}]}$, and so does X_t. The proof of the inequality below follows the guidelines of the proof of *i)*, using both Remark 2.2.4 and Corollary 2.2.12, which provide respectively the positive constants C_1', v_1 and C_2', v_2.

$$\begin{aligned}
W_1(X_t, \tilde{X}_t) &\leq \mathbb{E}\left[|X_t - \tilde{X}_t|\right] \\
&\leq \mathbb{E}\left[|X_t - \tilde{X}_t| \,\middle|\, \tau_A > t\right] \mathbb{P}(\tau_A > t) + \mathbb{E}\left[|X_t - \tilde{X}_t| \,\middle|\, \tau_A \leq t\right] \mathbb{P}(\tau_A \leq t) \\
&\leq \left((\mathbb{E}[X_0 + \tilde{X}_0]) \left(1 + \frac{1}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) + \frac{2\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]} \right) \mathbb{P}(\tau_A > t) \\
&\quad + \mathbb{E}\left[|X_t - \tilde{X}_t| \,\middle|\, \tau_A \leq t\right] \\
&\leq \left((\mathbb{E}[X_0 + \tilde{X}_0]) \left(1 + \frac{1}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) + \frac{2\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]} \right) \left(C_1' \mathrm{e}^{-v_1 t} + C_2' \mathrm{e}^{-v_2 t}\right).
\end{aligned}$$

It is easy to see that

$$\mathbb{E}\left[|\Theta_t - \tilde{\Theta}_t| \,\middle|\, \tau_A > t\right] \leq \mathbb{E}[\Theta_{N_t+1}] + \mathbb{E}[\tilde{\Theta}_{\tilde{N}_t+1}] \leq 2\mathbb{E}[\Theta],$$

and that

$$\mathbb{E}\left[\left|A_t - \tilde{A}_t\right| \,\middle|\, \tau_A > t\right] \leq \mathbb{E}[\Delta T_{N_t+1}] + \mathbb{E}[\tilde{\Delta}\tilde{T}_{\tilde{N}_t+1}] \leq 2\mathbb{E}[\Delta T].$$

Finally, we can conclude by writing that

$$\begin{aligned}
W_1(Y_t, \tilde{Y}_t) &\leq \mathbb{E}\left[|Y_t - \tilde{Y}_t| \,\middle|\, \tau_A > t\right] \mathbb{P}(\tau_A > t) + \mathbb{E}\left[|Y_t - \tilde{Y}_t| \,\middle|\, \tau_A \leq t\right] \mathbb{P}(\tau_A \leq t) \\
&\leq C_1 \mathrm{e}^{-v_1 t} + C_2 \mathrm{e}^{-v_2 t},
\end{aligned}$$

denoting by

$$C_1 = \left((\mathbb{E}[X_0 + \tilde{X}_0]) \left(1 + \frac{1}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) + \frac{2\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]} + 2\mathbb{E}[\Theta] + 2\mathbb{E}[\Delta T] \right) C_1',$$

and by

$$C_2 = \left((\mathbb{E}[X_0 + \tilde{X}_0]) \left(1 + \frac{1}{\mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]}\right) + \frac{2\mathbb{E}[U]}{1 - \mathbb{E}\left[\mathrm{e}^{-\Theta\Delta T}\right]} \right) C_2'.$$

$\square$

Remark 2.3.2: Proving the convergence in Wasserstein distance in (2.1.3) is quite easier than the convergence in total variation, and may still be improved by optimizing in α. Moreover, it does not require any assumption on F but a finite expectation, thus holds under assumptions (H2) and (H3) only. $\diamondsuit$

Note that we could also use a mixture of the Wasserstein distance for X and $\tilde{X}$, and the total variation distance for the second and third components, as in [BLBMZ12]; indeed, the processes Θ and $\tilde{\Theta}$ on the one hand, and A and $\tilde{A}$ on the other hand are interesting only when they are equal, i.e. when their distance in total variation is equal to 0.

2.3.2 Exponential inter-intake times

We turn to the particular case where $G = \mathscr{E}(\lambda)$ and f is Hölder with compact support, and we present another coupling method with a random division of the interval $[0, t]$. As highlighted above, the assumption on G is not relevant in a dietary context, but offers very simple and explicit rates of convergence. The assumption on f is pretty mild, given that this function represents the intakes of some chemical. It is possible, a priori, to deal easily with classical unbounded distributions the same way (like exponential or χ^2 distributions, provided that η is easily computable). We will not treat the convergence in Wasserstein distance (as in Theorem 2.1.1.*ii)*), since the mechanisms are roughly the same.

We provide two methods to bound the rate of convergence of the process in this particular case. On the one hand, the first method is a slight refinement of the speeds we got in Theorem 2.1.1, since the laws are explicit. On the other hand, we notice that the law of N_t is known and explicit calculations are possible. Thus, we do not split the interval $[0, t]$ into deterministic areas, but into random areas: $[0, T_1], [T_1, T_{N_t}], [T_{N_t}, t]$.

Firstly, let

$$\rho = 1 - \mathbb{E}\left[e^{-\Theta_1 T_1}\right].$$

Using the same arguments as in the proof of Lemma 2.2.15, one can easily see that

$$\sup_{x \in [0,\varepsilon]} \eta(x) \leq K \frac{M+1}{2} \varepsilon^h, \tag{2.3.2}$$

if $|f(x) - f(y)| \leq K|x - y|^h$ and $f(x) = 0$ for $x > M$.

Proposition 2.3.3

For $\alpha, \beta \in (0, 1), \alpha < \beta$,

$$\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq 1 - \left(1 - e^{-\lambda \alpha t}\right)\left(1 - e^{-\lambda(1-\beta)t}\right)\left(1 - Ce^{-\frac{\lambda \rho h}{1+h}(\beta-\alpha)t}\right)\left(1 - K\frac{M+1}{2} e^{-\frac{\lambda \rho h}{1+h}(\beta-\alpha)t}\right),$$

where $C = (\mathbb{E}[X_0 + \tilde{X}_0])\left(1 + \frac{1}{1-\rho}\right) + \frac{2\mathbb{E}[U]}{\rho}$.

We do not give the details of the proof because they are only slight refinements of the bounds in (2.3.1), with parameter $\varepsilon = \exp\left(-\frac{\lambda\rho(\beta-\alpha)}{1+h}t\right)$, since the rates of convergence are $v_2' = \lambda\rho$ and $v_4' = h$. This choice optimizes the speed of convergence, as highlighted in Remark 2.3.1. Note that the constant C could be improved since $\psi_{N_{\alpha t}}$ is known, but this is a detail which does not change the rate of convergence. Anyway, we can optimize these bounds by setting $\beta = 1-\alpha$ and $\alpha = \frac{\rho h}{1+h+2\rho h}$, so that the following inequality holds:

$$\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq 1 - \left(1 - \exp\left(\frac{-\lambda\rho h}{1+h+2\rho h}t\right)\right)^2 \left(1 - C\exp\left(\frac{-\lambda\rho h}{1+h+2\rho h}t\right)\right) \left(1 - K\frac{M+1}{2}\exp\left(\frac{-\lambda\rho h}{1+h+2\rho h}t\right)\right).$$

Then, developping the previous quantity, there exists $C_1 > 0$ such that

$$\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq C_1 \exp\left(\frac{-\lambda\rho h}{1+h+2\rho h}t\right). \tag{2.3.3}$$

Before exposing the second method, the following lemma is based on standard properties of the homogeneous Poisson processes, that we recall here.

Lemma 2.3.4

Let N be a homogeneous Poisson process of intensity λ.

i) $N_t \overset{\mathscr{L}}{=} \mathscr{P}(\lambda t)$.

ii) $\mathscr{L}(T_1, T_2, \ldots, T_n | N_t = n)$ *has a density* $(t_1, \ldots, t_n) \mapsto t^{-n} n! \mathbb{1}_{\{0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq t\}}$.

iii) $\mathscr{L}(T_1, T_n | N_t = n)$ *has a density* $g_n(u,v) = t^{-n} n(n-1)(v-u)^{n-2} \mathbb{1}_{\{0 \leq u \leq v \leq t\}}$.

Since $\mathscr{L}(T_1, T_n | N_t = n)$ is known, it is possible to provide explicit and better results in this specific case.

Proposition 2.3.5

For all $\varepsilon < 1$, the following inequality holds:

$$\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq 1 - \left(1 - e^{-\lambda t}\left(1 + \lambda t + \frac{\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon(1-\rho)^2}\left(e^{\lambda(1-\rho)t} - 1 - \lambda(1-\rho)t\right)\right)\right) \left(1 - K\frac{M+1}{2}\varepsilon^h\right).$$

Proof: Let $0 < \varepsilon < 1$ and $(Y, \tilde{Y})$ be the coupling generated by $\mathcal{L}_2$ in (2.2.8) between

0 and T_{N_t-1} and be the TV coupling between T_{N_t-1} and t. First, if $n \geq 2$, then

$$\begin{aligned}
\mathbb{P}\left(|X_{T_{N_t}^-} - \tilde{X}_{T_{N_t}^-}| \geq \varepsilon \middle| N_t = n \right) &\leq \frac{1}{\varepsilon}\mathbb{E}\left[|X_{T_{N_t}^-} - \tilde{X}_{T_{N_t}^-}| \middle| N_t = n \right] \\
&\leq \frac{1}{\varepsilon}\iint_{\mathbb{R}^2} \mathbb{E}\left[|X_{T_{N_t}^-} - \tilde{X}_{T_{N_t}^-}| \middle| N_t = n, T_1 = u, T_n = v \right] g_n(u,v)dudv \\
&\leq \frac{n(n-1)\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon t^n} \iint_{[0,t]^2} \mathrm{e}^{-\lambda\rho(v-u)}(v-u)^{n-2}\mathbb{1}_{\{u \leq v\}} dudv \\
&\leq \frac{n(n-1)\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon t^n} \int_0^t \mathrm{e}^{-\lambda\rho w}(t-w)w^{n-2}dw.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{P}\left(\left| X_{T_{N_t}^-} - \tilde{X}_{T_{N_t}^-} \right| \geq \varepsilon \right) &\leq \mathrm{e}^{-\lambda t}(1+\lambda t) \\
&\quad + \frac{\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon}\mathrm{e}^{-\lambda t}\sum_{n=2}^{\infty}\int_0^t \frac{\lambda^n}{(n-2)!}\mathrm{e}^{-\lambda\rho w}(t-w)w^{n-2}dw \\
&\leq \mathrm{e}^{-\lambda t}(1+\lambda t) + \frac{\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon}\lambda^2\mathrm{e}^{-\lambda t}\int_0^t \mathrm{e}^{-\lambda\rho w}\mathrm{e}^{\lambda w}(t-w)dw \\
&\leq \mathrm{e}^{-\lambda t}\left(1 + \lambda t + \frac{\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon(1-\rho)^2}\left(\mathrm{e}^{\lambda(1-\rho)t} - 1 - \lambda(1-\rho)t \right) \right).
\end{aligned}$$

Then, we use Proposition 2.2.14, Lemma 2.2.15 and (2.3.2) to conclude. □

Now, let us develop the inequality given in Proposition 2.3.5:

$$\begin{aligned}
\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq & K\frac{M+1}{2}\varepsilon^h + (1+\lambda t)\mathrm{e}^{-\lambda t} - K\frac{M+1}{2}(1+\lambda t)\mathrm{e}^{-\lambda t}\varepsilon^h \\
& + \frac{\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon(1-\rho)^2}\mathrm{e}^{-\lambda\rho t} - \frac{K(M+1)\mathbb{E}[X_0 \vee \tilde{X}_0]}{2\varepsilon(1-\rho)^2}\mathrm{e}^{-\lambda\rho t}\varepsilon^h \\
& - \frac{\mathbb{E}[X_0 \vee \tilde{X}_0]}{\varepsilon(1-\rho)^2}(1+\lambda(1-\rho)t)\mathrm{e}^{-\lambda t} \\
& - \frac{K(M+1)\mathbb{E}[X_0 \vee \tilde{X}_0]}{2\varepsilon(1-\rho)^2}(1+\lambda(1-\rho)t)\mathrm{e}^{-\lambda t}\varepsilon^h
\end{aligned}$$

The only fact that matters is that the first and the fourth terms in the previous expression are the slowest to converge to 0, thus it is straightforward that the rate of convergence is optimized by setting

$$\varepsilon = \exp\left(-\frac{\lambda\rho}{1+h}t \right),$$

and then there exists $C_2 > 0$ such that

$$\|\mu_0 P_t - \tilde{\mu}_0 P_t\|_{TV} \leq C_2 \exp\left(-\frac{\lambda\rho h}{1+h}t \right). \tag{2.3.4}$$

One can easily conclude, by comparing (2.3.3) and (2.3.4) that the second method provides a strictly better lower bound for the speed of convergence of the process to equilibrium.

CHAPTER 3

LONG TIME BEHAVIOR OF PIECEWISE DETERMINISTIC MARKOV PROCESSES

This chapter gathers various isolated results for Markov processes. They will allow us to link the work in this manuscript to other fields of research, such as stochastic approximation algorithms, Partial Differential Equation (PDE) analysis and shot-noise processes. Finally, in Section 3.3, we address the question of the link between the speed of mixing of a *Piecewise Deterministic Markov Process* (PDMP) and of its reversed-time version.

3.1 Convergence of a limit process for bandits algorithms

In this section, we study a PDMP called the *penalized bandit process*, whose dynamics were introduced by Lamberton and Pagès in [LP08b]. The behavior of this process is close to the one of the PDMPs that we studied in Chapter 2. We provided quantitative rates of convergence toward a stationary measure for these PDMPs in Theorem 2.1.1. Nonetheless, we relied deeply on the density of the jumps to get total variation convergence; here, the height of the jumps is deterministic, so we shall have to modify our coupling.

Note: this section is an adaptation of [BMP+15, Section 2]. The same results have been obtained in parallel by Gadat, Panloup and Saadane in the paper [GPS15].

3.1.1 The penalized bandit process

The *two-armed bandit algorithm* is a theoretical procedure to choose asymptotically the most profitable arm of a slot machine, or bandit; it was also used in the fields of mathematical psychology and of engineering. This algorithm has been widely studied, for instance in [LPT04, LP08a]. The key idea is to use a (deterministic) sequence of learning rates, rewarding an arm if it delivers a gain. Depending on the speed of convergence to 0 of this sequence, the algorithm is often faillible (it would not always select asymptotically the right arm, see [LPT04]).

It is possible to improve its results and ensure infaillibility by introducing penalties when the arm does not deliver a gain: this modification is called the *Penalized Bandit Algorithm* (PBA), and it is studied in [LP08b]. The authors show that, with a correct choice of penalties and rewards, and with the appropriate renormalization, the algorithm converges weakly to a probabiliy measure π, which is the stationary distribution of the PDMP with the following infinitesimal generator

$$\mathcal{L}f(x) = (1-p-px)f'(x) + qx\frac{f(x+g)-f(x)}{g},$$

where $0 < q < p < 1$, p and q being the respective probabilities of gain of the two arms. Surprisingly, the limit distribution is not Gaussian, as it could be expected since numerous stochastic approximation algorithms are ruled by a Central Limit Theorem (CLT) (see [KY03, For15]). The positive parameter g runs the asymptotic behaviour of the sequences of the rewards and penalties (see Section 3 in [LP08b] for details); for the sake of simplicity, we set $g = 1$ in the sequel. Moreover, the interval $[0,(1-p)/p)$ is transient, and computations are easier if we study the translated process $Y = X - \frac{1-p}{p}$, driven by the following generator:

$$\mathcal{L}^Y f(y) = -pyf'(y) + q\left(y + \frac{1-p}{p}\right)\big(f(y+1) - f(y)\big). \tag{3.1.1}$$

It is standard to deduce the dynamics of the process from the generator (see [Dav93]): between the jumps, Y satisfies $Y_t' = -pY_t$, and it jumps with jump rate $t \mapsto \zeta(Y_t) = q\left(Y_t + \frac{1-p}{p}\right)$ from Y_t to $Y_t + 1$.

In [LP08b], the authors show that π admits a density with support $[(1-p)/p, +\infty)$ and exponential moments of order up to u_M, where u_M is the unique positive solution of the equation

$$\frac{\mathrm{e}^{u_M} - 1}{u_M} = \frac{p}{q}. \tag{3.1.2}$$

Below, we recover (3.1.2) with a different argument (see Remark 3.1.2).

In the sequel, we call *penalized bandit process* the process Y of initial distribution μ_0 following the dynamics of $\mathcal{L}^Y$, and by μ_t its law at time t. Since the dynamics of this process are close to the ones of the pharmacokinetic PDMPs mentionned above, we turn to the study of its convergence to the stationary measure. As we will see, it is pretty easy to deduce Wasserstein exponential convergence from a very simple coupling argument, similar to the one used in [Bou15]: the key point is the monotonicity of such a coupling. However, we shall use another approach to show total variation convergence. Indeed, since the law of the gain is deterministic, we have to use the density of the jump times instead of the density of the jump height. But first, let us focus on bringing the two processes close to each other.

3.1.2 Wasserstein convergence

Firstly, recall the definitions of Wasserstein and the total variation distances between two measures on $\mathbb{R}$. Let $n \geq 1$:

$$W_n(\mu,\nu) = \inf\left\{\mathbb{E}[|X-Y|^n]^{\frac{1}{n}} : (X,Y) \text{ coupling of } \mu \text{ and } \nu\right\},$$
$$\|\mu-\nu\|_{TV} = \inf\left\{\mathbb{P}(X\neq Y) : (X,Y) \text{ coupling of } \mu \text{ and } \nu\right\}.$$

In the following, let μ_0 and $\widetilde{\mu}_0$ be two probabilities on $\mathbb{R}_+$ which admit a moment of order 1, and denote by μ_t (respectively $\widetilde{\mu}_t$) the law of the penalized bandit process at time t when its initial distribution is μ_0 (respectively $\widetilde{\mu}_0$). The following proposition holds:

Proposition 3.1.1

We have, for all $t \geq 0$

$$W_1(\mu_t,\widetilde{\mu}_t) \leq W_1(\mu_0,\widetilde{\mu}_0)\mathrm{e}^{-(p-q)t}. \tag{3.1.3}$$

Proof: Let $(Y,\widetilde{Y})$ be generated by

$$\begin{aligned}\mathcal{L}_2^Y f(y,\widetilde{y}) =& -py\partial_y f(y,\widetilde{y}) - p\widetilde{y}\partial_{\widetilde{y}} f(y,\widetilde{y}) + q(y-\widetilde{y})(f(y+1,\widetilde{y})-f(y,\widetilde{y}))\\ &+q\left(\widetilde{y}+\frac{1-p}{p}\right)(f(y+1,\widetilde{y}+1)-f(y,\widetilde{y})),\end{aligned} \tag{3.1.4}$$

for $y \geq \widetilde{y}$, and of symetric expression for $\widetilde{y} \geq y$, and such that $(Y_0,\widetilde{Y}_0)$ is a coupling of $(\mu_0,\widetilde{\mu}_0)$ realizing the equality $W_1(\mu_0,\widetilde{\mu}_0) = \mathbb{E}\left[|Y_0-\widetilde{Y}_0|\right]$. Taking $f(y,\widetilde{y}) = f(y)$, it is straightforward that Y follows the dynamics of $\mathcal{L}^Y$ in (3.1.1), and similarly for $\widetilde{Y}$, so that $(Y_t,\widetilde{Y}_t)_{t\geq 0}$ generated with $\mathcal{L}_2^Y$ is a coupling of $(\mu_t,\widetilde{\mu}_t)_{t\geq 0}$. With this coupling, either the higher process jumps alone or the two processes jump simultaneously. It is easy to check that this coupling is monotonous, *i.e.* for all $t\geq 0, (Y_t-\widetilde{Y}_t)(Y_0-\widetilde{Y}_0)\geq 0$. Monotonicity comes from the fact that the higher process jumps more often but stays above the other process since the jumps are positive. Assume that $\widetilde{Y}_0 \geq Y_0$. By monotonicity, we have, for all $t \geq 0$,

$$\mathbb{E}\left[|Y_t-\widetilde{Y}_t|\right] = \mathbb{E}[\widetilde{Y}_t]-\mathbb{E}[Y_t],$$

so that the proof reduces to the computation of $h : t \mapsto \mathbb{E}[Y_t]$. With $f(y) = y$, (3.1.1) leads to $\mathcal{L}f(y) = \frac{q(1-p)}{p} - (p-q)y$, and, by Dynkin's formula, the function h satisfies the ordinary differential equation $h'(t) = \frac{q(1-p)}{p} - (p-q)h(t)$. One deduces immediately that

$$\mathbb{E}[Y_t] = \frac{q(1-p)}{p(p-q)} + \left(\mathbb{E}[Y_0] - \frac{q(1-p)}{p(p-q)}\right) \mathrm{e}^{-(p-q)t}, \tag{3.1.5}$$

(recall that $p > q$). Then,

$$\mathbb{E}\left[|Y_t - \widetilde{Y}_t|\right] = \mathbb{E}\left[\widetilde{Y}_0 - Y_0\right] \mathrm{e}^{-(p-q)t},$$

which leads directly to (3.1.3) by definition of W_1 and the choice of $(Y_0, \widetilde{Y}_0)$. □

Remark 3.1.2: The Dynkin's formula is a powerful tool for studying the moments of Markov processes. One can use it with $f(y) = \mathrm{e}^{uy}$ to study the Laplace transform $\psi(\mu_t, u) = \mathbb{E}[\mathrm{e}^{uY_t}]$ of the process $(Y_t)_{t\geq 0}$. We have

$$\mathcal{L}^Y f(y) = q\frac{1-p}{p}(e^u - 1)f(y) + (q(e^u - 1) - up)yf(y),$$

so ψ satisfies the following PDE:

$$\partial_t \psi(\mu_t, u) = q\frac{1-p}{p}(e^u - 1)\psi(\mu_t, u) + (q(e^u - 1) - up)\partial_u \psi(\mu_t, u).$$

If $\mu_0 = \pi$, then $\partial_t \psi(\mu_t, u) = 0$, so that

$$\partial_u \left(\log(\psi(\pi, u))\right) = \frac{q\frac{1-p}{p}(\mathrm{e}^u - 1)}{up - q(\mathrm{e}^u - 1)},$$

and the right-hand side is finite for $u \in [0, u_M)$, when u_M is the solution of Equation (3.1.2). ♢

Note that the set of polynomials of degree n is stable under the action of $\mathcal{L}^Y$. This is an important property, since it theoretically enables us to compute the moments of Y_t by induction, with the help of Dynkin's formula, just as we did for the first moment in the proof of Proposition 3.1.1. Similarly, it is possible to study the function $h_n(t) = \mathbb{E}[|Y_t - \widetilde{Y}_t|^n]$, and then $W_n(\mu_t, \widetilde{\mu}_t)$. Indeed, we have, for $f(y, \widetilde{y}) = |y - \widetilde{y}|^n$,

$$\mathcal{L}_2^Y f(y, \widetilde{y}) = -n(p-q)|y - \widetilde{y}|^n + q\sum_{k=0}^{n-2} \binom{n}{k} |y - \widetilde{y}|^{k+1},$$

so that

$$h_n'(t) = -n(p-q)h_n(t) + q\sum_{k=0}^{n-2}\binom{n}{k} h_{k+1}(t).$$

Then, using Grönwall lemma, we derive by induction that $h_n(t) = O(\mathrm{e}^{-n(p-q)t})$. Which leads to the following result:

Proposition 3.1.3

For all $n \in \mathbb{N}^\star$, if μ_0 and $\widetilde{\mu}_0$ admit a moment of order n, there exists a constant

$C_n < +\infty$ such that, for all $t \geq 0$,

$$W_n(\mu_t, \widetilde{\mu}_t) \leq C_n e^{-(p-q)t}.$$

3.1.3 Total variation convergence

In the case of the penalized bandit process, total variation convergence is more challenging than in [Bou15], since the jumps are always of size 1. Instead, we are going to use the arguments introduced in [BCG+13b], based on the following observation: if Y and $\widetilde{Y}$ are close enough, we can make them jump, not simultaneously like before, but with a slight delay for one of the copies, which would make it jump on the other one, as illustrated in Figure 3.1.1.

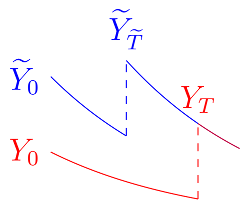


Figure 3.1.1 – Expected behaviour of the coalescent coupling for the penalized bandit process.

In the following, denote by $\tau = \inf\{t \geq 0 : \forall s \geq 0, Y_{t+s} = \widetilde{Y}_{t+s}\}$ the coalescence time of Y and $\widetilde{Y}$. The goal of the sequel is to obtain exponential moments for τ (which we expect for correct couplings) and then use the classic coupling inequality:

$$\|\mu_t - \widetilde{\mu}_t\|_{TV} \leq \mathbb{P}(Y_t \neq \widetilde{Y}_t) \leq \mathbb{P}(\tau > t).$$

We have the following lemma:

Lemma 3.1.4

Assume there exist positive constants $y_M < +\infty, \varepsilon_M < 1$ such that $Y_0, \widetilde{Y}_0 \leq y_M$ and $|Y_0 - \widetilde{Y}_0| \leq \varepsilon_M$. Then, there exist a coupling $(Y_t, \widetilde{Y}_t)_{t\geq 0}$ of $(\mu_t, \widetilde{\mu}_t)_{t\geq 0}$ and an explicit positive constant $C(y_M, \varepsilon_M) < +\infty$ such that, for all $t > 0$,

$$\mathbb{P}(\tau > t) \leq C(y_M, \varepsilon_M)\left(\exp\left(-\frac{q(1-p)}{p}t\right) + \varepsilon_M\right). \tag{3.1.6}$$

Proof: First, assume that Y_0 and $\widetilde{Y}_0$ are deterministic, and denote by $y = Y_0, \varepsilon = \widetilde{Y}_0 - y$. We assume w.l.o.g. that $\varepsilon > 0$. Let T (resp. $\widetilde{T}$) be the first jump time of the process Y (resp. $\widetilde{Y}$). Following the heuristics suggested by Figure 3.1.1, it is straightforward that

$$Y_T = \widetilde{Y}_T \Leftrightarrow T = \frac{1}{p}\log\left(\varepsilon + e^{p\widetilde{T}}\right).$$

Easy computations lead to

$$\mathbb{P}\left(\frac{1}{p}\log\left(\varepsilon+\mathrm{e}^{p\widetilde{T}}\right)\leq s\right)=\mathbb{P}\left(\widetilde{T}\leq\frac{1}{p}\log\left(\mathrm{e}^{ps}-\varepsilon\right)\right)=1-\Phi_y(s,\varepsilon),$$

with

$$\Phi_y(s,\varepsilon)=\exp\left(-\frac{q}{p}\left(\frac{1-p}{p}\log(\mathrm{e}^{ps}-\varepsilon)+(y+\varepsilon)\left(1-\frac{1}{\mathrm{e}^{ps}-\varepsilon}\right)\right)\right).$$

As a consequence, the random variables T and $\frac{1}{p}\log(\varepsilon+\exp(p\widetilde{T}))$ admit densities w.r.t. the Lebesgue measure, which are respectively $f_y(\cdot,0)$ and $f_y(\cdot,\varepsilon)$, with, for all $s\geq 0$,

$$f_y(s,\varepsilon)=\frac{q\mathrm{e}^{ps}}{p}\left(\frac{1-p}{\mathrm{e}^{ps}-\varepsilon}+\frac{p(y+\varepsilon)}{(\mathrm{e}^{ps}-\varepsilon)^2}\right)\Phi_y(s,\varepsilon).$$

Let T and $\frac{1}{p}\log(\varepsilon+\exp(p\widetilde{T}))$ follow the so-called γ-coupling, (the coupling minimizing the total variation of their laws, see [Lin92]). It is not hard to deduce from the very construction of this coupling the following equality:

$$\mathbb{P}\left(T=\frac{1}{p}\log\left(\varepsilon+\mathrm{e}^{p\widetilde{T}}\right),T<t\right)=\int_0^t\left(f_y(s,0)\wedge f_y(s,\varepsilon)\right)ds,$$

and then

$$\begin{aligned}\mathbb{P}\left(\tau\leq t\right)=\mathbb{P}\left(Y_t=\widetilde{Y}_t\right)&\geq\mathbb{P}\left(T=\frac{1}{p}\log\left(\varepsilon+\mathrm{e}^{p\widetilde{T}}\right),T<t\right)\\&\geq 1-\frac{1}{2}\left(\Phi_y(t,0)+\Phi_y(t,\varepsilon)+\int_0^t|f_y(s,0)-f_y(s,\varepsilon)|ds\right).\end{aligned}\tag{3.1.7}$$

From the definition of Φ_y, the following upper bound is easily obtained for any $0\leq\varepsilon\leq\varepsilon_M$ and any $0\leq y\leq y_M$:

$$\Phi_y(s,\varepsilon)\leq C_1\exp\left(-\frac{q(1-p)}{p}s\right),\tag{3.1.8}$$

with $C_1=\exp\left(\frac{-q}{p}\left(\frac{1-p}{p}\log(1-\varepsilon_M)+(y_M+\varepsilon_M)\left(1-\frac{1}{1-\varepsilon_M}\right)\right)\right)$. In order to apply the mean-value theorem, we differentiate f_y with respect to ε. After some computations, one can obtain the following upper bound:

$$\left|\frac{\partial f_y}{\partial\varepsilon}(s,\varepsilon)\right|\leq C_1C_2\exp\left(-\frac{q(1-p)}{p}s\right),\tag{3.1.9}$$

where

$$\begin{aligned}C_2=&\frac{q^2((1-p)(1-\varepsilon_M)+p(y_M+\varepsilon_M))}{p^2(1-\varepsilon_M)^2}\left(1\vee\left(\frac{(2-p)(1-\varepsilon_M)+y_M+\varepsilon_M}{(1-\varepsilon_M)^2}\right)\right)\\&+\frac{q(1-\varepsilon_M+2p(y_M+\varepsilon_M))}{p(1-\varepsilon_M)^3}.\end{aligned}$$

Then, we easily have

$$\int_0^t |f_y(s,0) - f_y(s,\varepsilon)| ds \leq C_1 C_2 \varepsilon \int_0^{+\infty} \exp\left(-\frac{q(1-p)}{p}s\right) ds \leq \frac{pC_1C_2}{q(1-p)}\varepsilon_M. \tag{3.1.10}$$

Combining Equations (3.1.7), (3.1.8), (3.1.9) and (3.1.10) entails (3.1.6) with

$$C(y_M, \varepsilon_M) = C_1 + \frac{pC_1C_2}{2q(1-p)}. \tag{3.1.11}$$

The upper bound provided in (3.1.11) does not depend on Y_0 and $\widetilde{Y}_0$, so this result still holds for random starting points, provided that they belong to $[0, y_M]$. □

Proposition 3.1.1 and Lemma 3.1.4 are the main tools to prove exponential convergence in total variation:

Proposition 3.1.5

Let $t_0 > 0$. There exists an explicit positive constant $K < +\infty$ (see (3.1.14)) such that, for all $t \geq t_0$,

$$\|\mu_t - \widetilde{\mu}_t\|_{TV} \leq K \exp\left(-\frac{p-q}{2 + \frac{p(p-q)}{q(1-p)}}t\right). \tag{3.1.12}$$

Proof: Let $\alpha \in (0,1)$ and $u > 0$. We use first the coupling from Proposition 3.1.1 in the domain $[0, \alpha t]$ to bring the processes close to each other and next the coupling from Lemma 3.1.4 in the domain $[\alpha t, t]$ to bring them to coalescence. We set $\varepsilon_M = \mathrm{e}^{-ut}$ and $y_M = \frac{q(1-p)}{p(p-q)} + 1$, and use the following inequality:

$$\begin{aligned} \mathbb{P}(\tau \leq t) \geq \mathbb{P}\left(|Y_{\alpha t} - \widetilde{Y}_{\alpha t}| \leq \varepsilon_M, Y_{\alpha t} \vee \widetilde{Y}_{\alpha t} \leq y_M\right) \\ \times \mathbb{P}\left(\tau \leq t \,\middle|\, |Y_{\alpha t} - \widetilde{Y}_{\alpha t}| \leq \varepsilon_M, Y_{\alpha t} \vee \widetilde{Y}_{\alpha t} \leq y_M\right). \end{aligned} \tag{3.1.13}$$

On the one hand, (3.1.5) leads to

$$\begin{aligned} \mathbb{P}\left(Y_{\alpha t} \geq y_M\right) &= \mathbb{P}\left(Y_{\alpha t} - \frac{q(1-p)}{p(p-q)} \geq y_M - \frac{q(1-p)}{p(p-q)}\right) \leq \mathbb{P}\left(\left|Y_{\alpha t} - \frac{q(1-p)}{p(p-q)}\right| \geq 1\right) \\ &\leq \left|\mathbb{E}\left[Y_0 - \frac{q(1-p)}{p(p-q)}\right]\right| \exp(-\alpha(p-q)t), \end{aligned}$$

so that

$$\begin{aligned} \mathbb{P}\left(|Y_{\alpha t} - \widetilde{Y}_{\alpha t}| > \varepsilon_M \text{ or } Y_{\alpha t} \vee \widetilde{Y}_{\alpha t} > y_M\right) &\leq \mathbb{P}\left(|Y_{\alpha t} - \widetilde{Y}_{\alpha t}| > \varepsilon_M\right) + \mathbb{P}\left(Y_{\alpha t} > y_M\right) \\ &\quad + \mathbb{P}\left(\widetilde{Y}_{\alpha t} > y_M\right) \\ &\leq C_3 \exp((u - \alpha(p-q))t), \end{aligned}$$

with $C_3 = \left(W_1(\mu_0,\widetilde{\mu}_0) + \left|\mathbb{E}\left[Y_0 - \frac{q(1-p)}{p(p-q)}\right]\right| + \left|\mathbb{E}\left[\widetilde{Y}_0 - \frac{q(1-p)}{p(p-q)}\right]\right|\right)$. On the other hand, let $C_4 = \sup_{t\geq t_0} C(y_M, \mathrm{e}^{-ut})$. The constant C_4 is finite and, from Lemma 3.1.4,

$$\mathbb{P}\left(\tau > t \middle| |Y_{\alpha t} - \widetilde{Y}_{\alpha t}| \leq \varepsilon_M, Y_{\alpha t} \vee \widetilde{Y}_{\alpha t} \leq y_M\right) \leq C_4\left(\mathrm{e}^{-ut} + \exp\left(-\frac{q(1-p)(1-\alpha)}{p}t\right)\right).$$

Now, (3.1.13) reduces to

$$\begin{aligned}\mathbb{P}(\tau > t) \leq 1 - (1 - C_3 \exp((u - \alpha(p-q))t))\\ \times\left(1 - C_4\left(\mathrm{e}^{-ut} + \exp\left(-\frac{q(1-p)(1-\alpha)}{p}t\right)\right)\right).\end{aligned}$$

Optimizing the rate of convergence by setting

$$\alpha = \frac{1}{1 + \frac{p(p-q)}{2q(1-p)}}, \qquad u = \frac{(p-q)\alpha}{2} = \frac{p-q}{2 + \frac{p(p-q)}{q(1-p)}},$$

we derive (3.1.12) with

$$K = C_3 + 2C_4. \tag{3.1.14}$$

□

3.2 Links with other fields of research

3.2.1 Growth/fragmentation equations and processes

We consider the growth and division of a population of micro-organisms (typically, bacteria or cells) through a quantity x which rules the division. For instance, one can consider x to be the size of the bacterium. We refer to [Per07, Chapter 4] for background and biological motivations, and to [DJG10] and the references therein for motivations for determining eigenelements of a PDE, and for a wide range of other applications. In the article [CDG12], the authors investigate the behavior of the Malthusian parameter (or fitness) of the population. This coefficient is the first eigenvalue of a PDE, and the authors study its dependence on the growth and division rates. The aim of this section is to go over the aforementioned article from a probabilistic point of view, to explain the assumptions for the well-posedness of the problem and to draw connections between probability theory and PDE theory. We provide a probabilistic justification to the links between the growth and fragmentation rates, with the help of the renowned Foster-Lyapunov criterion.

The evolution of the population, or rather its probabilistic interpretation, has also been studied, in the context of network congestions, as the *Transmission Control Protocol* (TCP) window size process (see [LvL08, CMP10, ABG+14, DHKR15]). This phenomenon yields to the following PDE, for $x, t \geq 0$:

$$\partial_t u(t,x) + \partial_x[\tau(x)u(t,x)] + \beta(x)u(t,x) = 2\int_x^\infty \beta(y)\kappa(x,y)u(t,y)dy, \tag{3.2.1}$$

with boundary conditions $u(0,x) = u_0(x), u(t,0) = 0$. In the literature, such an equation is referred to as *growth/fragmentation equation*. The quantity $u(t,x)$ represents the concentration of individuals of size x at time t. The size of a bacterium grows at rate τ, i.e. following the differential equation $\partial_t y = \tau(y)$. The bacteria of size x break into two daughters at rate $\beta(x)$, following a fragmentation kernel $\kappa(y,x)$, which is the proportion of bacteria of size y born from a mother with size x. The factor 2 in the right-hand side of (3.2.1) represents the binary division of a mother into two daughters. Since κ represents a proportion, we assume that, for $x > 0$,

$$\int_0^x \kappa(y,x)dy = 1, \tag{3.2.2}$$

Many biological models would require $\int_0^x y\kappa(y,x)dy = x/2$, so that the mass of the mother is conserved after the fragmentation, which is automatically satisfied for a symmetric division $\kappa(y,x) = \kappa(x-y,x)$.

Let us provide a probabilistic interpretation of this mechanism. Consider a bacterium of size X, which grows at rate τ and randomly splits at rate β following a kernel κ, as before. For the sake of coherence with the models studied in this manuscript, we denote by $Q(x,dy) = x\kappa(yx,x)dy$, so that

$$\int_0^x f(y)\kappa(y,x)dy = \int_0^1 f(xy)Q(x,dy),$$

to deal with the size of the daughters compared to the size of the mother. From (3.2.2), we have, for any $x > 0$, $\int_0^1 Q(x,dy) = 1$. If we dismiss one of the two daughters and carry on the study with the other one, this phenomenon can also be classically (as explained in Chapter 1) modeled with a PDMP $(X_t)_{t\geq 0}$ generated by

$$\mathcal{L}f(x) = \tau(x)f'(x) + \beta(x)\int_0^1 [f(xy) - f(x)]Q(x,dy). \tag{3.2.3}$$

We shall call $(X_t)_{t\geq 0}$ a *growth/fragmentation process*, which is Feller (see [Dav93]), and we denote by (P_t) its semigroup. As mentionned in Section 1.2.4, if we denote by $\mu_t = \mathscr{L}(X_t)$, the Kolmogorov's forward equation $\partial_t(P_t f) = \mathcal{L}P_t f$ is the weak formulation of

$$\partial_t \mu_t = \mathcal{L}'\mu_t, \tag{3.2.4}$$

where $\mathcal{L}'$ is the adjoint operator of $\mathcal{L}$ in $L^2(\mathbb{L})$ and $\mathbb{L}$ is the Lebesgue measure. Note that $\mathcal{L}'$ is different from $\mathcal{L}^*$ defined in Section 3.3, which is the adjoint of $\mathcal{L}$ in $L^2(\pi)$, π being the invariant measure of X. If μ_t admits a density $u(t,\cdot)$ over $\mathbb{R}_+$, then (3.2.4) writes

$$\partial_t u(t,x) + \partial_x[\tau(x)u(t,x)] + \beta(x)u(t,x) = \int_x^\infty \beta(y)\kappa(x,y)u(t,y)dy, \tag{3.2.5}$$

with boundary conditions $u(0,x) = u_0(x), u(t,0) = 0$. Note that (3.2.5) is the conservative version of (3.2.1), since for any $t \geq 0$, $\int u(t,x)dx = 1$, which comes from the fact that there is only one bacterium at a time.

To investigate the assumptions used in [CDG12], we turn to the study of the Markov process generated by (3.2.3). More precisely, we will provide a justification to the balance between τ and β assumed in (2.4) and (2.5) with the help of a Foster-Lyapunov criterion. Note that we shall not require the fragmentation kernel $Q(x, dy)$ to admit a density with respect to the Lebesgue measure $\mathbb{L}(dy)$. Moreover, in order to be as general as possible, we do not stick to the biological framework and thus do not assume that $\int_0^1 Q(x, dy) = 1/2$, which will be replaced by Assumption 3.2.2.i). Now, let us make general assumptions on the growth and fragmentation rates.

Assumption 3.2.1 (*Behavior of τ and β*)

i) The functions β and τ are continuous, and τ is locally Lipschitz.

ii) For any $x > 0$, $\beta(x), \tau(x) > 0$.

iii) There exist constants $\gamma_0, \gamma_\infty, \nu_0, \nu_\infty \in \mathbb{R}$ and $\beta_0, \beta_\infty, \tau_0, \tau_\infty > 0$ such that

$$\beta(x) \underset{x\to 0}{\sim} \beta_0 x^{\gamma_0}, \quad \beta(x) \underset{x\to\infty}{\sim} \beta_\infty x^{\gamma_\infty}, \quad \tau(x) \underset{x\to 0}{\sim} \tau_0 x^{\nu_0}, \quad \tau(x) \underset{x\to\infty}{\sim} \tau_\infty x^{\nu_\infty}.$$

Note that Assumption 3.2.1.iii) is purely technical, and is not required for the ergodicity to hold (see Assumptions (2.21) and (2.22) in [CDG12]). If τ and β satisfy Assumption 3.2.1, then Assumptions (2.18) and (2.19) in [CDG12] are fulfilled (by taking $\mu = |\gamma_\infty|$ or $\mu = |\nu_\infty|$, and $r_0 = |\nu_0|$).

The following assumption concerns the expected behavior of the fragmentation, and is easy to check in most cases, especially if $Q(x, \cdot)$ does not depend on x. For any $a \in \mathbb{R}$, we define the moment of order a of $Q(x, \cdot)$:

$$M_x(a) = \int_0^1 y^a Q(x, dy).$$

Assumption 3.2.2 (*Moments of Q*)

i) There exist $a > 0$ such that $\sup_{x>0} M_x(a) < 1$.

ii) There exist $b > 0$ such that $\sup_{x>0} M_x(-b) < +\infty$.

Note that, in particular, Assumption 3.2.2 implies that, for any $x > 0$,

$$Q(x, \{1\}) = Q(x, \{0\}) = 0.$$

We can now state the main result of this section.

Proposition 3.2.3 (*Stability of growth/fragmentation processes*)

Let X be the PDMP generated by (3.2.3). If Assumption 3.2.1 holds, then X is irreducible and aperiodic, and compact sets are petite. Moreover, if Assumption 3.2.2 holds, and if

$$\gamma_0 + 1 - \nu_0 > 0, \quad \gamma_\infty + 1 - \nu_\infty > 0,$$

then the process X possesses a unique stationary measure π. Furthermore, if

$$\gamma_\infty \geq 0, \quad \nu_0 \leq 1,$$

then X is exponentially ergodic.

Remark 3.2.4 (*Use of a Lyapunov function in the analysis of the PDE*): Note that Assumption 3.2.2 is sufficient but not necessary to deduce ergodicity from a Foster-Lyapunov criterion, since we only need the limits in (3.2.9) and (3.2.10) to be negative. Namely, we ask the fragmentation kernel not to be too close to 0 and 1. Regardless, the goal is to find a and b as large as possible, so that we have a Lyapunov function defined in (3.2.8) as coercive as possible. Indeed, if Theorem 1.2.4 holds, then $V \in L^1(\pi)$. Even if the stationary measure is not explicit, determining its moments is usually a good beginning to understand the behavior of a Markov process; see for example [LvL08, Section 3] and [BCG13a]. For many growth/fragmentation processes, it is possible to build a Lyapunov function of the form $x \mapsto \mathrm{e}^{\omega x}$, thus π admits exponential moments up to ω. Incidentally, we use a close approach and the existence of the Laplace transform in the proof of Proposition 3.3.4. $\diamondsuit$

Proof of Proposition 3.2.3: Firstly, let us prove that compact sets are petite for $(X_t)_{t\geq 0}$. We shall denote by φ_z the unique maximal solution of $\partial_t y(t) = \tau(y(t))$ with initial condition z. Let $z_2 > z_1 > z_0 > 0$ and $z \in [z_0, z_1]$. Since $\tau > 0$ on $[z_0, z_2]$, the function φ_z is a diffeomorphism from $[0, \varphi_z^{-1}(z_2)]$ to $[z, z_2]$; let $t = \varphi_{z_0}^{-1}(z_2)$ be the maximum time for the flow to reach z_2 from $[z_0, z_1]$. Denote by X^z the process generated by (3.2.3) such that $\mathscr{L}(X_0) = \delta_z$, and T_n^z the epoch of its n^{th} jump. Let $\mathscr{A} = \mathscr{U}([0,t])$. For any $x \in [z_1, z_2]$, we have

$$\begin{aligned}
\int_0^\infty \mathbb{P}(X_s^z \leq x)\mathscr{A}(ds) &\geq \frac{1}{t}\int_0^t \mathbb{P}(X_s^z \leq x | T_1^z > \varphi_z^{-1}(z_2))\mathbb{P}(T_1^z > \varphi_z^{-1}(z_2))ds \\
&\geq \frac{\mathbb{P}(T_1^z > \varphi_z^{-1}(z_2))}{t}\int_0^t \mathbb{P}(\varphi_z(s) \leq x)ds \\
&\geq \frac{\mathbb{P}(T_1^z > \varphi_z^{-1}(z_2))}{t}\int_z^x (\varphi_z^{-1})'(u)du. \qquad (3.2.6)
\end{aligned}$$

Since β and τ are bounded on $[z_0, z_2]$, the following inequalities hold:

$$\begin{aligned}
\mathbb{P}(T_1^z > \varphi_z^{-1}(z_2)) &= \exp\left(-\int_0^{\varphi_z^{-1}(z_2)} \beta(\varphi_z(s))ds\right) = \exp\left(-\int_z^{z_2} \beta(u)(\varphi_z^{-1})'(u)du\right) \\
&\geq \exp\left(-(z_2 - z_0)\sup_{[z_0,z_2]} \beta(\varphi_z^{-1})'\right) \\
&\geq \exp\left(-(z_2 - z_0)\left(\sup_{[z_0,z_2]}\beta\right)\left(\inf_{[z_0,z_2]}\tau\right)^{-1}\right),
\end{aligned}$$

$$\inf_{[z_0,z_2]}(\varphi_z^{-1})' = \left(\sup_{[z_0,z_2]}\tau\right)^{-1}.$$

Hence, there exists a constant C such that, (3.2.6) writes, for $x \in [z_1, z_2]$,

$$\int_0^\infty \mathbb{P}(X_s^z \leq x)\mathscr{A}(ds) \geq C(x - z_1),$$

which rewrites

$$\int_0^\infty \delta_z P_s \mathscr{A}(ds) \geq C\mathbb{L}_{[z_1,z_2]},$$

where $\mathbb{L}_K$ is the Lebesgue measure restricted to a set K. Hence, by definition, $[z_0, z_1]$ is a petite set for X.

Now, let us show that the process (X_t) is $\mathbb{L}_{(0,\infty)}$-irreducible. Let $z_1 > z_0 > 0$ and $z > 0$. If $z \leq z_0$,

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{z_0 \leq X_t^z \leq z_1\}} dt\right] &\geq \mathbb{P}(T_1^z > \varphi_z^{-1}(z_1))\mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{z_0 \leq X_t^z \leq z_1\}} dt \middle| T_1^z > \varphi_z^{-1}(z_1)\right] \\ &\geq \exp\left(-(z_1 - z_0)\left(\sup_{[z_0,z_1]} \beta\right)\left(\inf_{[z_0,z_1]} \tau\right)^{-1}\right)\varphi_{z_0}^{-1}(z_1). \quad (3.2.7)\end{aligned}$$

If $z > z_0$, for any $t_0 > 0$ and $n \in \mathbb{N}$, the process X^z has a positive probability of jumping n times before time t_0. Denote by $p = \sup_{x>0} M_x(a)$. For any $n > (\log(z) - \log(z_0))\log(p^{-1})^{-1}$, let $0 < \varepsilon < z_0^a - (zp^n)^a$. By continuity of $(z, t) \mapsto \varphi_z(t)$ and since $\int_0^1 y^a Q(x, dy) \leq p < 1$, there exists $t_0 > 0$ small enough such that

$$\mathbb{E}[(X_{t_0}^z)^a | T_n^z \leq t_0] \leq (zp^n)^a + \varepsilon < z_0^a, \quad \mathbb{P}(X_{t_0}^z \leq z_0 | T_n^z \leq t_0) \geq 1 - \frac{\mathbb{E}[(X_{t_0}^z)^a | T_n^z \leq t_0]}{z_0^a} > 0.$$

Then, $\mathbb{P}(X_{t_0}^z \leq z_0) > 0$ for any t_0 small enough, and, using (3.2.7)

$$\begin{aligned}\mathbb{E}\left[\int_0^\infty \mathbb{1}_{\{z_0 \leq X_t^z \leq z_1\}} dt\right] &\geq \mathbb{E}\left[\int_{t_0}^\infty \mathbb{1}_{\{z_0 \leq X_t^z \leq z_1\}} dt \middle| X_{t_0}^z \leq z_0\right]\mathbb{P}(X_{t_0}^z \leq z_0) \\ &\geq \exp\left(-(z_1 - z_0)\left(\sup_{[z_0,z_1]} \beta\right)\left(\inf_{[z_0,z_1]} \tau\right)^{-1}\right)\varphi_{z_0}^{-1}(z_1)\mathbb{P}(X_{t_0}^z \leq z_0) \\ &> 0.\end{aligned}$$

Aperiodicity is easily proven with similar arguments.

Let $a, b > 0$ be as defined in Assumption 3.2.2, and let V be a smooth, convex function on $(0, \infty)$ defined by

$$V(x) = \begin{cases} x^{-b} & \text{if } x \in (0, 1], \\ x^a & \text{if } x \in [2, \infty). \end{cases} \quad (3.2.8)$$

For $x \geq 2$, $V(x) = x^a$ and

$$\begin{aligned}
\mathcal{L}V(x) &= a\frac{\tau(x)}{x}V(x) + \beta(x)\int_0^1 V(xy)Q(x,dy) - \beta(x)V(x) \\
&\leq \left(a\frac{\tau(x)}{x} - \beta(x)\right)V(x) + \beta(x)\int_0^{1/x}(xy)^{-b}Q(x,dy) \\
&\quad + \beta(x)\int_{1/x}^{2/x} 2^a Q(x,dy) + \beta(x)\int_{2/x}^{1}(xy)^a Q(x,dy) \\
&\leq \left(a\frac{\tau(x)}{x} - \beta(x)\right)V(x) + \beta(x)\left(x^{-b}M_x(-b) + 2^a + x^a M_x(a)\right) \\
&\leq \left(a\frac{\tau(x)}{x} - \beta(x)\left(1 - M_x(a) - \frac{M_x(b)}{x^bV(x)} - \frac{2^a}{V(x)}\right)\right)V(x).
\end{aligned}$$

Combining $\gamma_\infty + 1 - \nu_\infty > 0$ with Assumption 3.2.2,

$$\begin{aligned}
a\frac{\tau(x)}{x} &- \beta(x)\left(1 - M_x(a) - \frac{M_x(b)}{x^bV(x)} - \frac{2^a}{xV(x)}\right) \\
&\leq a\frac{\tau(x)}{x} - \beta(x)\left(1 - \sup_{x>0} M_x(a) + o(1)\right) \leq 0
\end{aligned}$$

for x large enough. For $x \leq 1$, $V(x) = x^{-b}$ and

$$\mathcal{L}V(x) = \left(-b\frac{\tau(x)}{x} + \beta(x)(M_x(-b) - 1)\right)V(x).$$

Likewise, combining $\gamma_0 + 1 - \nu_0 > 0$ with Assumption 3.2.2.ii),

$$-b\frac{\tau(x)}{x} + \beta(x)(M_x(-b) - 1) \leq -b\frac{\tau(x)}{x} + \beta(x)\left(\sup_{x>0} M_x(-b) - 1\right) \leq 0$$

for x close enough to 0. Then, [MT93b, Theorem 3.2] shows that X is Harris recurrent, thus admits a unique stationary measure (see for instance [KM94]).

Now, if we assume $\gamma_\infty \geq 0$ and $\nu_0 \leq 1$ in addition, then there exists $\alpha > 0$ such that

$$\begin{aligned}
\lim_{x\to+\infty} a\frac{\tau(x)}{x} &- \beta(x)\left(1 - M_x(a) - \frac{M_x(b)}{x^bV(x)} - \frac{2^a}{xV(x)}\right) \\
&\leq \lim_{x\to+\infty} a\frac{\tau(x)}{x} - \beta(x)\left(1 - \sup_{x>0} M_x(a) + o(1)\right) \leq -\alpha,
\end{aligned} \tag{3.2.9}$$

and

$$\lim_{x\to 0} -b\frac{\tau(x)}{x} + \beta(x)(M_x(-b) - 1) \leq \lim_{x\to 0} -b\frac{\tau(x)}{x} + \beta(x)\left(\sup_{x>0} M_x(-b) - 1\right) \leq -\alpha. \tag{3.2.10}$$

Combining (3.2.9) and (3.2.10), and since V is bounded on $[1,2]$, there exist positive constants $A, \alpha' > 0$ such that

$$\mathcal{L}V \leq -\alpha V + \alpha' \mathbb{1}_{[1/A, A]}.$$

The function V is a Lyapunov function satisfying the assumptions of Theorem 1.2.4, which applies and achieves the proof. □

3.2.2 Shot-noise decomposition of piecewise deterministic Markov processes

In this section, we shall show how we can write a PDMP as a shot-noise process. The literature about shot-noise processes is very rich, and we refer to [Ric77], as well as [HT89] and the references therein for some examples of the topics in which shot-noise processes arise. There are slightly different ways of defining them, so we shall follow [IJ03], and say that a shot-noise process is a stochastic process $(X_t)_{t\geq 0}$ in $\mathbb{R}^d$ which admits a general decomposition of the form

$$X_t = \sum_{n=0}^{\infty} g_n(t - T_n),$$

where the T_n are the epochs of a (possibly delayed) renewal process and the g_n are stochastic processes with right continuous with left limits (càdlàg) trajectories almost surely (a.s.) We call renewal process the *backward recurrence time process* defined in [Asm03, Chapter 5], which is the time elapsed since the last epoch. For $n \geq 1$, the random variables $T_{n+1} - T_n$ are independent and identically distributed (i.i.d.), and the process is delayed whenever $\mathscr{L}(T_1) \neq \mathscr{L}(T_2 - T_1)$. The term $g_n(t - T_n)$ can be interpreted as the effect at time t of an event, occurring at time T_n with a random effect g_n characterizing the event (magnitude, type, etc.). A particular case of this decomposition is when

$$g_n(t - T_n) = g(t - T_n)U_n,$$

where U_n is a sequence of random vectors and g is a deterministic càdlàg function. Following [BD12], which deals with one-dimensional shot-noise processes, we call U_n the *impulse* of the n^{th} event, and g the *kernel function* of the shot-noise, which characterizes the way the events are felt. For instance, the case $g(t) = \mathrm{e}^{-t}\mathbb{1}_{t\geq 0}$ has been widely studied (see among others [OB83, IJ03, BD12]) and we will see that it is strongly linked to the pharmacokinetic process introduced in Remark 1.1.1.

The shot-noise processes have already been intensively studied, but considering them as PDMPs could lead to new breakthroughs thanks to the rich literature about PDMPs. Conversely, linking PDMPs to shot-noise processes might be interesting in many areas:

- As we briefly mentionned in Chapter 1, level crossings are of particular interest in the domain of statistics. This has already been studied in the setting of shot-noise processes in [OB83, BD12].
- Results of regularity for the law of shot-noise processes have already been proven in [OB83, Bre10] for instance.
- The long time behavior of shot-noise processes as been deeply studied, as well as their stationary distributions or the limit theorems they satisfy; see for instance [IJ03] or [Iks13, IMM14].

Proposition 3.2.5 (*Shot-noise processes and PDMPs*)

Let $(X_t)_{t\geq 0}$ be a stochastic process on $\mathbb{R}^d$, and $M \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$. The two following *statements are equivalent:*

i) The process $(X_t)_{t\geq 0}$ is a shot-noise process with decomposition

$$\forall t \geq 0, \quad X_t = \sum_{n=0}^{\infty} g_n(t - T_n), \tag{3.2.11}$$

with g_0 the unique solution of $\partial_t y = My + b$, $g_n(t) = e^{tM} U_n \mathbb{1}_{t\geq 0}$ for $n \geq 1$ and $(U_n)_{n\geq 1}$ is a sequence of i.i.d. random vectors.

ii) There exists a renewal process $(A_t)_{t\geq 0}$ such that $(X_t, A_t)_{t\geq 0}$ is a PDMP with infinitesimal generator

$$\mathcal{L}f(x,a) = (Mx+b)\nabla_x f(x,a) + \partial_a f(x,a) + \zeta(a) \int_{\mathbb{R}^d} [f(x+u, 0) - f(x,a)] Q(du), \tag{3.2.12}$$

with $Q \in \mathscr{M}_1$ and $\zeta : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{+\infty\}$.

Whenever these statements hold, $\mathscr{L}(U_n) = Q$. Moreover, the T_n are the epochs of (A_t) and ζ is the hazard rate of $\mathscr{L}(T_{n+1} - T_n)$.

We refer to [Bon95] for deeper insights about reliability and hazard rates. Note that, necessarily, $g_0(0) = X_0$ and that the present notation is coherent with the one of Chapter 2. In fact, Proposition 3.2.5 captures the class of PDMPs studied in Chapter 2 as soon as there exists $\theta > 0$ such that $H = \delta_\theta$; in other words, when the metabolic parameter is constant the pharmacokinetic process may be written as a shot-noise process. With the decomposition provided in Proposition 3.2.5, X_t can be seen as the effect of every jump which occured at time $T_n \leq t$, which can not be felt before the jump since $g_n(t) = 0$ if $t \leq 0$. If we set $d = 1, M = -\theta, b = 0, Q = \mathscr{E}(\alpha)$, we recover the pharmacokinetic process, and we can see the quantity of contaminant at time t as the cumulated sum of the remaining contaminant ingested at time $T_n < t$, namely

$$X_t = X_0 \mathrm{e}^{-\theta t} + \sum_{n=1}^{\infty} U_n \mathrm{e}^{-\theta(t-T_n)} \mathbb{1}_{T_n \leq t}.$$

Remark 3.2.6 (*Interpretation of Proposition 3.2.5*): With only a linear vector field, the framework of (3.2.12) may seem restrictive at first glance, but it captures several PDMPs mentionned in Chapters 1, 2 and 3, which are used when modeling natural phenomena: among others, the TCP window-size process and the pharmacokinetic process of Remark 1.1.1. As a matter of fact, it is hard to hope for more general PDMPs to admit a shot-noise decomposition. For instance, PDMPs with switching, as studied in [FGM12, BLBMZ14, BL14] can not fit in our framework, since for shot-noise processes, the effect of a jump is always felt the same way (i.e. with the same kernel function) after its occurrence. Switching from $\partial_t y = My + b$ to $\partial_t y = M'y + b'$ would require to change the influence of all the previous jumps, or to include correcting terms into U_{n+1} taking into account $X_0, U_1, \ldots, U_n$. and the previous drift terms. ♢

Proof of Proposition 3.2.5: Firstly, let $N_t = \sup\{n \in \mathbb{N} : T_n \leq t\}$ and φ be the unique solution of $\partial_t y = My + b$ with initial condition 0 (we have $\varphi(t) = (\mathrm{e}^{tM} - I_d)M^{-1}b$ if M is invertible). Then, $\Phi(x,t) = \varphi(t) + \mathrm{e}^{tM}x$ is the unique solution of $\partial_t y = My + b$ with initial condition x, and, by setting $U_0 = X_0$,

$$\sum_{n=0}^{\infty} g_n(t - T_n) = \varphi(t) + \sum_{n=0}^{N_t} \mathrm{e}^{(t-T_n)M} U_n = \Phi\left(\sum_{n=0}^{N_t} \mathrm{e}^{-T_n M} U_n, t\right). \tag{3.2.13}$$

The proof of *ii)*⇒*i)* is based on a simple recursion. Denote by T_n the jump times of (X, A). Obviously, $X_t = g_0(t)$ if $t < T_1$. Now assume that, for some $n \geq 1$ and every $s \in [0, T_n)$, $X_s = \sum_{k=0}^{n-1} g_k(s - T_k)$. Let $t \in [T_n, T_{n+1})$ and $U_n \sim Q$. We have

$$\begin{aligned}
X_t &= \Phi\left(\Phi(X_{T_{n-1}}, \Delta T_n) + U_n, t - T_n\right) = \mathrm{e}^{(t-T_n)M}(\Phi(X_{T_{n-1}}, \Delta T_n) + U_n) + \varphi(t - T_n) \\
&= \Phi\left(\Phi(X_{T_{n-1}}, \Delta T_n), t - T_n\right) + \mathrm{e}^{(t-T_n)M} U_n = \Phi(X_{T_{n-1}}, t - T_{n-1}) + \mathrm{e}^{(t-T_n)M} U_n \\
&= \Phi\left(\sum_{k=0}^{n-1} \mathrm{e}^{(T_{n-1}-T_k)M} U_k, t - T_{n-1}\right) + \mathrm{e}^{(t-T_n)M} U_n \\
&= \varphi(t - T_{n-1}) + \mathrm{e}^{(t-T_{n-1})M} \varphi(T_{n-1}) + \sum_{k=0}^{n} \mathrm{e}^{(t-T_k)M} U_k = \varphi(t) + \sum_{k=0}^{n} \mathrm{e}^{(t-T_k)M} U_k.
\end{aligned} \tag{3.2.14}$$

Now, we turn to the proof of *i)*⇒*ii)*. For $t \geq 0$, let $A_t = t - T_{N_t}$; by definition, A is a renewal process with epochs T_n. Then, the stochastic process (X, A) admits càdlàg trajectories a.s. and, following the proof of [Asm03, Proposition 1.5, Chapter V], it is a strong Markov process. Now, combining (3.2.13) and (3.2.14), it is clear that $(X_t, A_t)_{t \geq 0}$ is generated by $\mathcal{L}$: X follows the flow $\partial_t y = My + b$, A follows the flow $\partial_t y = y$ and the process jumps at rate ζ from $(X_{T_n^-}, A_{T_n^-})$ to $(X_{T_n} + U_n, 0)$. □

3.3 Time-reversal of piecewise deterministic Markov processes

In this section, we turn to the study of the time-reversal of a stochastic process $(X_t)_{t \geq 0}$. Informally, it is the process having the dynamics of X when the times goes backward. If X is a stationary Markov process, we can define its time-reversal as the stochastic process $(X_t^*)_{t \geq 0}$ defined by

$$X_t^* = X_{(T-t)^-}$$

for some $T \geq 0$ (or a suitably defined random time). A natural goal is to relate the speeds of convergence to equilibrium of X and X^*. Unfortunately, this is presently beyond our reach, and in the following we bring out the main issues when we addressed this question, in the framework of two different PDMPs. We refer to [LP13b], which provides motivations for time-reversal, as well as a general method to compute X^* and its characteristics. For PDMPs with a discrete component, the reader can also check [FGR09]. The framework of this article includes most of the PDMPs presented in Chapters 1, 2 and 3, as well as PDMPs with switching.

In fact, it is always difficult to obtain quantitative speeds of convergence for PDMPs if the flow does not draw the trajectories together, as for growth/fragmentation processes. But whenever a flow is divergent, its opposite is convergent, and it might be easier to obtain speeds of mixing with this new flow. That is why linking the speed of convergence of a PDMP to the one of its time-reversal is of interest. And comparing Figures 3.3.3 and 3.3.5, we can reasonably assume that the speed of convergence to equilibrium for some PDMPs is the same than the one of their time-reversed version. Nevertheless, it is possible to compute the jump mechanism of the reversed process only when the stationary measure is tractable (see Lemmas 3.3.2 and 3.3.5), which is a strong motivation to get rates of convergence in the most general setting.

3.3.1 Reversed on/off process

We begin with the study of a simple PDMP with switching, called *on/off process* in [BKKP05]. Let $(Y_t)_{t\geq 0} = (X_t, I_t)_{t\geq 0}$ be the PDMP evolving on $\mathbb{Y} = (0,1) \times \{0,1\}$, driven by the infinitesimal generator:

$$\mathcal{L}f(x,i) = -\theta(x-i)\partial_x f(x,i) + \lambda\left[f(x,1-i) - f(x,i)\right], \tag{3.3.1}$$

for $\lambda > 0, \theta > 0, (x,i) \in \mathbb{Y}$. The process X continuously switches from one flow to the other, each of them exponentially attracting it toward 0 or 1 (see Figure 3.3.1). Similar switching processes can also be interpreted within a context of pharmacokinetics, where X represents the quantity of contaminant and I the current phase (*intake* or *assimilation*). Then, the class of PDMPs introduced in Chapter 2 may be interpreted as a limit process, if the time-scale of the intake is much shorter than the one of the assimilation. Similar two-scale phenomena may appear in gene expression models with bursting transcription (see [YZLM14]).

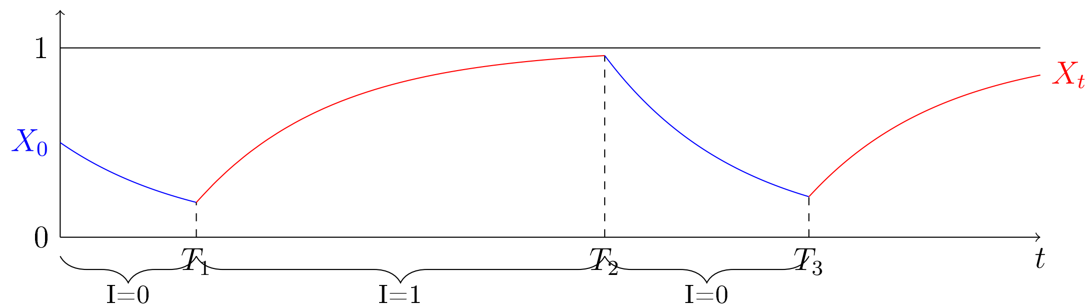


Figure 3.3.1 – Typical trajectory of the on/off process generated by $\mathcal{L}$ in (3.3.1).

Proposition 3.3.1

The Markov process $(Y_t)_{t\geq 0}$ generated by $\mathcal{L}$ in (3.3.1) admits a unique stationary measure on $\mathbb{Y}$

$$\pi = C_{\lambda,\theta}(\pi_0\otimes\delta_0+\pi_1\otimes\delta_1), \quad \pi_0(dx) = x^{\lambda/\theta-1}(1-x)^{\lambda/\theta}dx, \quad \pi_1(dx) = x^{\lambda/\theta}(1-x)^{\lambda/\theta-1}dx,$$

where $C_{\lambda,\theta} = \frac{1}{2}\beta(\lambda/\theta + 1, \lambda/\theta)^{-1}$. *Moreover,*

$$W_1(Y_t, \pi) \le \begin{cases} \left(2 + \frac{\lambda}{|\theta-\lambda|}\right)\exp(-\min(\lambda,\theta)t) & \text{if } \lambda \neq \theta \\ (2+\lambda t)e^{-\theta t} & \text{if } \lambda = \theta \\ W_1(Y_0,\pi)e^{-\theta t} & \text{if } \mathscr{L}(I_0) = \frac{1}{2}(\delta_0 + \delta_1) \end{cases}.$$

Proof: Using [BKKP05, Theorem 1], it is easy to check that the expression given for π entails $\pi(\mathcal{L}f) = 0$ for f smooth, thus π is a stationary measure for Y. Convergence toward equilibrium, as proved afterwards, ensures us of the uniqueness of π.

Now, we turn to the quantification of the ergodicity of the process. Since the flow is exponentially contracting, at rate θ, one can expect the Wasserstein distance of the spatial component X to decrease exponentially. The only problem is to bring I_t to its stationary measure first. So, consider the Markov process on $\mathbb{Y} \times \mathbb{Y}$ with infinitesimal generator

$$\begin{aligned} \mathcal{L}_2 f(x,i,\widetilde{x},\widetilde{i}) &= -\theta\left[(x-i)\partial_x + (\widetilde{x}-\widetilde{i})\partial_{\widetilde{x}}\right] f(x,i,\widetilde{x},\widetilde{i}) \\ &+ \lambda\left[f(x,1-i,\widetilde{x},1-\widetilde{i}) - f(x,i,\widetilde{x},\widetilde{i})\right]\mathbb{1}_{i=\widetilde{i}} \\ &+ \lambda\left[f(x,1-i,\widetilde{x},\widetilde{i}) - f(x,i,\widetilde{x},\widetilde{i})\right]\mathbb{1}_{i\neq\widetilde{i}} \\ &+ \lambda\left[f(x,i,\widetilde{x},1-\widetilde{i}) - f(x,i,\widetilde{x},\widetilde{i})\right]\mathbb{1}_{i\neq\widetilde{i}}. \end{aligned} \tag{3.3.2}$$

The coupling $(Y,\widetilde{Y}) = (X,I,\widetilde{X},\widetilde{I})$ generated by $\mathcal{L}_2$ in (3.3.2) evolves independently until $I = \widetilde{I}$, and with common flow and jumps afterwards. We set $T_0 = 0$ and denote by T_n the epoch of the n^{th} jump ; then, $T_{n+1} - T_n \sim \mathscr{E}(\lambda)$. If $I_0 \neq \widetilde{I}_0$, the first jump is a.s. not common, and then $I_{T_1} = \widetilde{I}_{T_1}$. Consequently,

$$\begin{aligned} \mathbb{E}\left[|Y_t - \widetilde{Y}_t|\right] &= \mathbb{E}\left[|X_t - \widetilde{X}_t|\right] + \mathbb{P}(I_t \neq \widetilde{I}_t) \\ &\le \int_0^t \mathbb{E}\left[|X_t - \widetilde{X}_t|\,\middle|\, T_1 = u\right]\lambda e^{-\lambda u}du + \int_t^\infty 2\lambda e^{-\lambda u}du \\ &\le 2e^{-\lambda t} + \int_0^t \mathbb{E}\left[|X_u - \widetilde{X}_u|\right]e^{-\theta(t-u)}\lambda e^{-\lambda u}du \le 2e^{-\lambda t} + \lambda e^{-\theta t}\int_0^t e^{(\theta-\lambda)u}du \\ &\le \left[\left(2 + \frac{\lambda}{\theta-\lambda}\right)e^{-\lambda t} - \frac{\lambda}{\theta-\lambda}e^{-\theta t}\right]\mathbb{1}_{\{\theta\neq\lambda\}} + (2+\lambda t)e^{-\lambda t}\mathbb{1}_{\{\theta=\lambda\}} \\ &\le \left(2 + \frac{\lambda}{|\theta-\lambda|}\right)e^{-(\theta\wedge\lambda)t}\mathbb{1}_{\{\theta\neq\lambda\}} + (2+\lambda t)e^{-\lambda t}\mathbb{1}_{\{\theta=\lambda\}} \end{aligned}$$

Finally, if $\mathscr{L}(I_0) = \frac{1}{2}(\delta_0 + \delta_1)$, the coupling $(Y,\widetilde{Y})$ always has common jumps and

$$|Y_t - \widetilde{Y}_t| = |Y_0 - \widetilde{Y}_0|e^{-\theta t},$$

and letting $(X_0, \widetilde{X}_0)$ be the optimal Wasserstein coupling is enough to ensure Wasserstein contraction. $\square$

Since the inter-jump times are spread-out, it is also possible to show convergence in total variation with a method similar to Proposition 3.1.5. But what about the reversed process? Since π is explicit, it is possible to compute the characteristics of the reversed process Y^*.

Lemma 3.3.2

Let Y be a PDMP generated by $\mathcal{L}$ in (3.3.1). Then, $Y^* = (X^*, I^*)$ is also a PDMP, with infinitesimal generator

$$\mathcal{L}^* f(x,i) = \theta(x-i)\partial_x f(x,i) + \lambda \frac{i-x}{x+i-1}[f(x,1-i) - f(x,i)]. \tag{3.3.3}$$

The characteristics of the reversed process Y^* generated by (3.3.3) are the following. When $I = i$, the flow $\partial_t y = \theta(x-i)$ drives X^* exponentially fast toward $(1-i)$, but the jump rate tends to $+\infty$ and the process switches to the other flow before hitting $(1-i)$: the dynamics of X^* are the very opposite of the ones of X. Of course, π is still a stationary measure for Y^*.

Proof of Lemma 3.3.2: Using [LP13b, Theorem 2.4], X^* is a PDMP evolving on $\mathbb{Y}$, with some infinitesimal generator denoted by

$$\mathcal{L}^* f(x,i) = F^*(x,i)\partial_x f(x,i) + \lambda^*(x,i) \int_{\mathbb{Y}} [f(y) - f(x,i)] Q^*\left((x,i), dy\right).$$

Firstly, since the deterministic dynamics between the jumps are reversed, we have to set $F^*(x,i) = \theta(x-i)$. Now, we use [LP13b, Theorem 2.4] to get the relation, for $y, y' \in \mathbb{Y}$,

$$\lambda Q(y, dy')\pi(dy) = \lambda^*(y')Q^*(y', dy)\pi(dy'), \tag{3.3.4}$$

where $Q((x,i), dy') = \delta_{(x,1-i)}(dy')$ is the jump kernel of the regular process. From the left-hand side of (3.3.4), the only possible choice for the jump kernel of the reversed process is

$$Q^*((x',i'), dy) = \delta_{(x',1-i')}(dy).$$

Then, for $(x,i) \in \mathbb{Y}$, (3.3.4) writes,

$$\lambda(x,i)\pi(d(x,i)) = \lambda\pi(d(x,1-i)).$$

Hence,

$$\lambda(x,0) = \lambda\frac{x}{1-x}, \quad \lambda(x,1) = \lambda\frac{1-x}{x}.$$

□

It is rather hard to obtain explicit speeds of convergence for the Wasserstein distance using coupling methods for the reversed process Y^*. Indeed, because of the exponential flow, two trajectories will not remain close to each other whatever the coupling we use. Total variation couplings are theoretically more easy to set up, but until now I did not obtain any conclusive result. Anyway, the useful Foster-Lyapunov criterion applies here and allows us to prove geometric ergodicity for Y^* (see Proposition 3.3.3 below). For hints about the real speeds of convergence in W_1, which seem similar for Y and Y^*, the reader may refer to Figures 3.3.2 and 3.3.3. For other results of ergodicity for switching processes, we refer to [BLBMZ15, CH15].

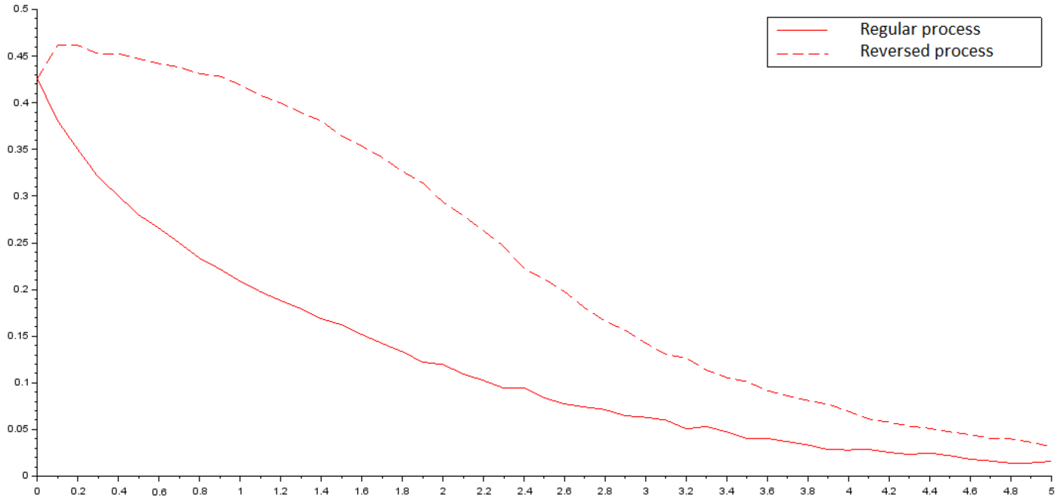


Figure 3.3.2 – Simulations of $t \mapsto W_1(Y_t, \pi)$ and $t \mapsto W_1(Y_t^*, \pi)$, for $\mathscr{L}(Y_0) = \mathscr{L}(Y_0^*) = \delta_{0.9} \otimes \delta_0, \theta = 1, \lambda = 0.5$.

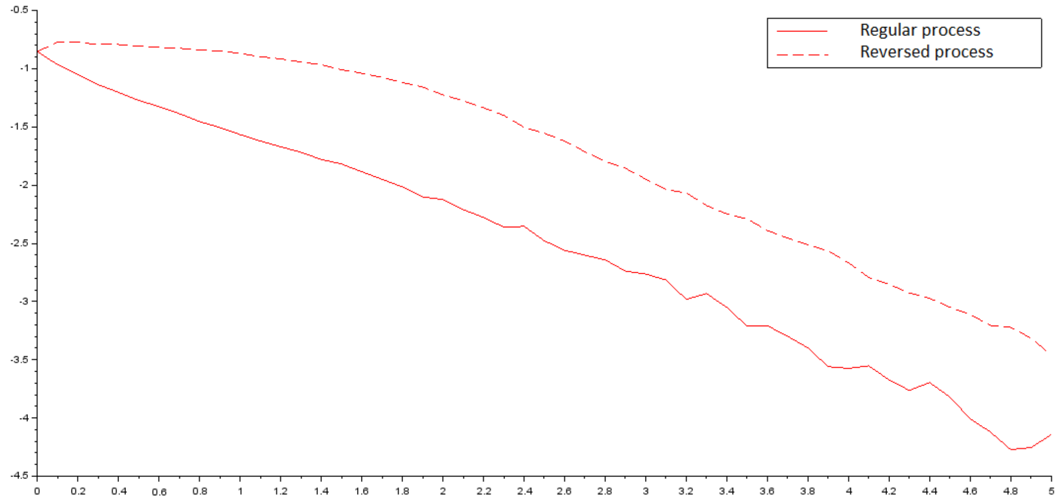


Figure 3.3.3 – Simulations of $t \mapsto \log(W_1(Y_t, \pi))$ and $t \mapsto \log(W_1(Y_t^*, \pi))$, for $\mathscr{L}(Y_0) = \mathscr{L}(Y_0^*) = \delta_{0.9} \otimes \delta_0, \theta = 1, \lambda = 0.5$.

Proposition 3.3.3 (***Geometric ergodicity of the reversed on/off process***)

For $(x, i) \in \mathbb{Y}$. Let $V : \mathbb{Y} \to (0, +\infty)$ and $\gamma \in (0, 1)$ such that

$$V(x, i) = x^\gamma (1-x)^{\gamma-1} \mathbb{1}_{i=0} + x^{\gamma-1}(1-x)^\gamma \mathbb{1}_{i=1}, \quad \gamma > 1 - \frac{\lambda}{\theta}.$$

There exist $C, v > 0$ such that, if $\mu = \mathscr{L}(Y_0^*) \in L^1(V)$,

$$\|Y_t^* - \pi\|_{TV} \leq C\mu(V)e^{-vt}.$$

Proof: The proof is a mere application of the Foster-Lyapunov criterion. Indeed, for $i = 0$:

$$\mathcal{L}^* V(x,0) = \left(\beta' - \alpha' \frac{x}{1-x}\right) V(x,0), \quad \alpha' = \lambda - \theta(1-\gamma), \quad \beta' = \lambda + \gamma\theta.$$

Since $\alpha' > 0$, we have, for any $a \in (0,1)$,

$$\beta' - \alpha' \frac{x}{1-x} \leq \begin{cases} \beta' & \text{if } 0 < x \leq a \\ -\alpha & \text{if } a < x < 1 \end{cases}, \quad \alpha = \frac{a\alpha'}{1-a} - \beta'.$$

Note that $\alpha > 0$ as soon as $a > \beta'(\alpha' + \beta')^{-1}$, and, for $\beta = (\alpha + \beta') \sup_{[0,a]} V(\cdot, 0)$,

$$\mathcal{L}^* V(x,0) \leq -\alpha V(x,0) + \beta \mathbb{1}_{[0,a]}(x).$$

Similar computations for $\mathcal{L}^* V(x,1)$ entail

$$\mathcal{L}^* V(x,i) \leq -\alpha V(x,i) + \beta \mathbb{1}_K(x),$$

where $K = [0,a] \times \{0\} \cup [1-a, 1] \times \{1\}$ is a compact of $\mathbb{Y}$.

It is straightforward but tedious to show that compact sets of $\mathbb{Y}$ are petite for $(X_t)_{t \geq 0}$, and that the process is irreducible and aperiodic. Computations are similar to the proof of Proposition 3.2.3. Then, Theorem 1.2.4 achieves the proof. □

3.3.2 Time-reversal in pharmacokinetics

In this section, we provide another example of time-reversed process, namely the pharmacokinetic process introduced in Remark 1.1.1. Let us consider a Markov process with infinitesimal generator:

$$\mathcal{L}f(x) = -\theta x f'(x) + \lambda \int_0^\infty [f(x+y) - f(x)] \alpha \mathrm{e}^{-\alpha y} dy. \tag{3.3.5}$$

Between its jumps, the process follows the flow given by the ODE $\partial_t X_t = -\theta X_t$, and jumps at times T_n, such that

$$\Delta T_n = T_n - T_{n-1}, \quad \Delta T_n \sim \mathscr{E}(\lambda), \quad \mathscr{L}(X_{T_n} - X_{T_n^-}) = \mathscr{E}(\alpha).$$

Proposition 3.3.4

The Markov process $(X_t)_{t \geq 0}$ generated by $\mathcal{L}$ in (3.3.5) admits a unique stationary measure $\pi = \Gamma(\lambda/\theta, 1/\alpha)$ on $\mathbb{R}_+^*$, with density

$$\pi(dx) = \frac{(\alpha x)^{\lambda/\theta - 1}}{\Gamma(\lambda/\theta)} \alpha e^{-\alpha x} dx.$$

Moreover,

$$W_1(X_t, \pi) \leq W_1(X_0, \pi) \mathrm{e}^{-\theta t}.$$

Proof: Existence and uniqueness of π are the result of a slight generalization [Mal15, Lemma 2.1]. Define $f(x) = \mathrm{e}^{ux}$ for $u \in (0, \alpha)$. We have

$$\begin{aligned}\mathcal{L}f(x) &= -\theta u x \mathrm{e}^{ux} + \lambda \alpha \mathrm{e}^{ux} \int_0^\infty (\mathrm{e}^{(u-\alpha)y} - \mathrm{e}^{-\alpha y}) dy \\ &= -\theta u (x \mathrm{e}^{ux}) + \frac{\lambda u}{\alpha - u} \mathrm{e}^{ux}.\end{aligned}$$

Let $\mathscr{L}(X_0)$ be some probability distribution with exponential moments up to α. Using Dynkin's formula, letting $\psi(u,t) = \mathbb{E}[\mathrm{e}^{uX_t}]$, we have,

$$\partial_t \psi(u,t) = -\theta u \partial_u \psi(u,t) + \frac{\lambda u}{\alpha - u} \psi(u,t).$$

Letting $t \to +\infty$, the Laplace transform of π satisfies the following ODE, for $0 < u < \alpha$,

$$0 = -\theta u \partial_u \psi_\pi(u) + \frac{\lambda u}{\alpha - u} \psi_\pi(u).$$

Simple computations provide the existence of a constant C such that

$$\psi_\pi(u) = C(\alpha - u)^{-\lambda/\theta} = \left(1 - \frac{u}{\alpha}\right)^{-\lambda/\theta}.$$

Then, one can easily conclude that $\pi = \Gamma(\lambda/\theta, 1/\alpha)$ is the only stationary measure for X.

Now, we turn to the study of the geometric ergodicity of X. As in Remark 1.2.6, consider the Markov process $(X, \widetilde{X})$ generated by

$$\mathcal{L}_2 f(x, \widetilde{x}) = -\theta \partial_x f(x, \widetilde{x}) - \theta \partial_{\widetilde{x}} f(x, \widetilde{x}) + \lambda \int_0^\infty [f(x+u, \widetilde{x}+u) - f(x, \widetilde{x})] \alpha \mathrm{e}^{-\alpha u} du,$$

and $(X_0, \widetilde{X}_0)$ being the optimal coupling in Wasserstein for $\mathscr{L}(X_0)$ and $\mathscr{L}(\widetilde{X}_0)$. The processes X and $\widetilde{X}$ follow the same flow and jump at the same time, so that

$$W_1(X_t, \widetilde{X}_t) \leq \mathbb{E}[|X_t - \widetilde{X}_t|] = W_1(X_0, \widetilde{X}_0) \mathrm{e}^{-\theta t}.$$

Let $\mathscr{L}(\widetilde{X}_0) = \pi$ to achieve the proof. $\square$

Since the stationary measure π is now explicit, it is rather simple to obtain the characteristics of the reversed process.

Lemma 3.3.5

Let X be a PDMP generated by $\mathcal{L}$ in (3.3.5). Then, X^* is also a PDMP, with infinitesimal generator

$$\mathcal{L}^* f(x) = \theta x f'(x) + \alpha \theta x \int_0^1 [f(xy) - f(x)] \frac{\lambda}{\theta} y^{\lambda/\theta - 1} dy. \tag{3.3.6}$$

The proof of this lemma is a mere application of [LP13b, Theorem 2.4]. Then, the behavior of X^* is depicted in Lemma 3.3.5: X^* is a growth/fragmentation process as introduced in Section 3.2.1, growing with the flow $\partial_t y(t) = \theta y(t)$ and jumping with a jump rate $\beta(x) = \alpha\theta x$, following the fragmentation kernel $Q(x, \cdot) = \beta(\lambda/\theta, 1)$. Note that Q does not depend on x, and that the process $X^\star$ satisfies Assumption 3.2.1 with $\nu_0 = \nu_\infty = \gamma_0 = \gamma_\infty = 1$. Moreover, under the notation of Assumption 3.2.2, for any $x > 0$, $M_x(1) = \lambda(\lambda+\theta)^{-1}$ and $\sup_{x\leq 1} M_x(-b) < +\infty$ as soon as $b < \lambda/\theta$. Thus, Proposition 3.2.3 entails the geometric ergodicity of X^*. We simulated the speeds of mixing of the processes X and X^* in Figures 3.3.4 and 3.3.5.

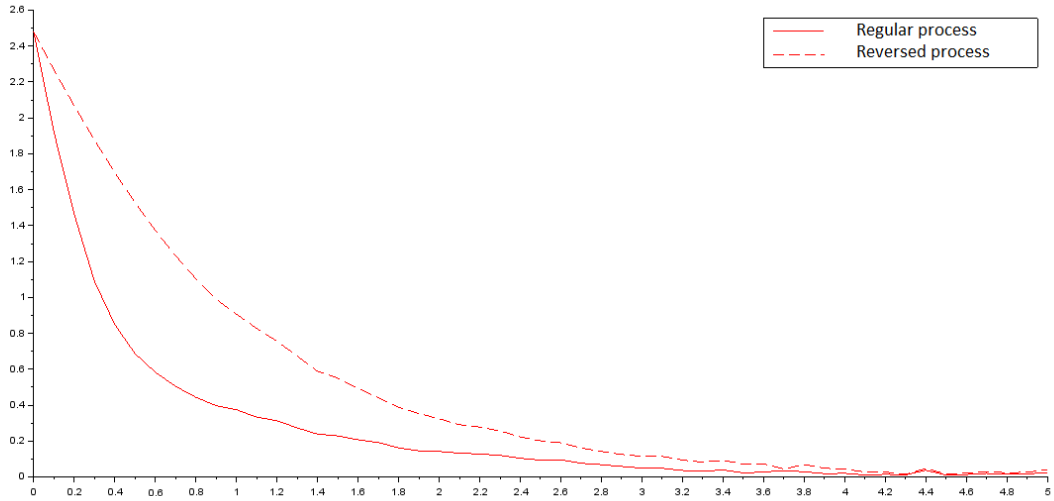


Figure 3.3.4 – Simulations of $t \mapsto W_1(X_t, \pi)$ and $t \mapsto W_1(X_t^*, \pi)$, for $\mathscr{L}(Y_0) = \mathscr{L}(Y_0^*) = \delta_5, \theta = 1, \lambda = 2, \alpha = 1/2$.

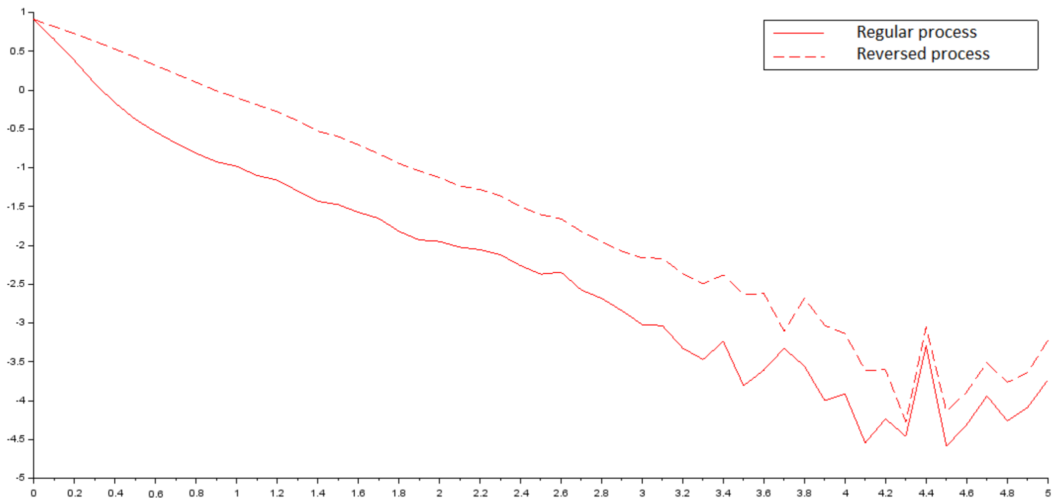


Figure 3.3.5 – Simulations of $t \mapsto \log(W_1(X_t, \pi))$ and $t \mapsto \log(W_1(X_t^*, \pi))$, for $\mathscr{L}(Y_0) = \mathscr{L}(Y_0^*) = \delta_5, \theta = 1, \lambda = 2, \alpha = 1/2$.

CHAPTER 4

STUDY OF INHOMOGENEOUS MARKOV CHAINS WITH ASYMPTOTIC PSEUDOTRAJECTORIES

In this chapter, we consider an inhomogeneous (discrete time) Markov chain and are interested in its long time behavior. We provide sufficient conditions to ensure that some of its asymptotic properties can be related to the ones of a homogeneous (continuous time) Markov process. Renowned examples such as a bandit algorithms, weighted random walks or decreasing step Euler schemes are included in our framework. Our results are related to functional limit theorems, but the approach differs from the standard "Tightness/Identification" argument; our method is unified and based on the notion of pseudotrajectories on the space of probability measures.

Note: this chapter is an adaptation of [BBC16].

4.1 Introduction

In this paper, we consider an inhomogeneous Markov chain $(y_n)_{n\geq 0}$ on $\mathbb{R}^D$, and a non-increasing sequence $(\gamma_n)_{n\geq 1}$ converging to 0, such that $\sum_{n=1}^{\infty} \gamma_n = +\infty$. For any smooth function f, we set

$$\mathcal{L}_n f(y) := \frac{\mathbb{E}\left[f(y_{n+1}) - f(y_n)|y_n = y\right]}{\gamma_{n+1}}. \tag{4.1.1}$$

We shall establish general asymptotic results when $\mathcal{L}_n$ converges, in some sense explained below, toward some infinitesimal generator $\mathcal{L}$. We prove that, under reasonable hypotheses, one can deduce properties (trajectories, ergodicity, etc) of $(y_n)_{n\geq 1}$ from the ones of a process generated by $\mathcal{L}$.

This work is mainly motivated by the study of the rescaling of stochastic approximation algorithms (see e.g. [Ben99, LP13a]). Classically, such rescaled algorithms converge to Normal distributions (or linear diffusion processes); see e.g. [Duf96, KY03, For15]. This Central Limit Theorem (CLT) is usually proved with the help of "Tightness/Identification" methods. With the same structure of proof, Lamberton and Pagès get a different limit in [LP08b]; namely, they provide a convergence to the stationary measure of a non-diffusive Markov process. Closely related, the decreasing step Euler scheme (as developed in [LP02, Lem05]) behaves in the same way.

In contrast to this classical approach, we rely on the notion of asymptotic pseudotrajectories introduced in [BH96]. Therefore, we focus on the asymptotic behavior of $\mathcal{L}_n$ using Taylor expansions to deduce immediately the form of a limit generator $\mathcal{L}$. A natural way to understand the asymptotic behavior of $(y_n)_{n\geq 0}$ is to consider it as an approximation of a Markov process generated by $\mathcal{L}$. Then, provided that the limit Markov process is ergodic and that we can estimate its speed of convergence toward the stationary measure, it is natural to deduce convergence and explicit speeds of convergence of $(y_n)_{n\geq 0}$ toward equilibrium. Our point of view can be related to the Trotter-Kato theorem (see e.g. [Kal02]). The proof of our main theorem, Theorem 4.2.7 below, is related to Lindeberg's proof of the CLT; namely it is based on a telescopic sum and a Taylor expansion.

With the help of Theorem 4.2.7, the study of the long time behavior of $(y_n)_{n\geq 0}$ reduces to the one of a homogeneous-time Markov process. Their convergence has been widely studied in the litterature, and we can differentiate several approaches. For instance, there are so-called "Meyn-and-Tweedie" methods (or Foster-Lyapunov criteria, see [MT93b, HM11, HMS11, CH15]) which provide qualitative convergence under mild conditions; we can follow this approach to provide qualitative properties for our inhomogeneous Markov chain. However, the speed is usually not explicit or very poor. Another approach consists in the use of *ad hoc* coupling methods (see e.g. [Lin92, Ebe11, Bou15]) either for a diffusion or a *Piecewise Deterministic Markov Process* (PDMP). Those methods usually prove themselves to be efficient for providing explicit speeds of convergence, but rely on extremely particular strategies. Among other approaches, let us also mention functional inequalities or spectral gap methods (see e.g. [Bak94, ABC$^+$00, Clo12, Mon14a]).

In this article, we develop a unified approach to study the long time behavior of inhomogeneous Markov chains, which may also provide speeds of convergence or functional convergence. To our knowledge, this method is original, and Theorems 4.2.7 and 4.2.9 have the advantage of being self-contained. The main goal of our illustrations, in Section 4.3, is to provide a simple framework to understand our approach. For these examples, proofs seem more simple and intuitive, and we are able to recover classical results as well as slight improvements.

This paper is organized as follows. In Section 4.2, we state the framework and the main assumptions that will be used throughout the paper. We recall the notion of asymptotic pseudotrajectory, and present our main result, Theorem 4.2.7, which describes the asymptotic behavior of a Markov chain. We also provide two consequences, Theorems 4.2.9 and 4.2.13, precising the geometric ergodicity of the chain or its functional convergence. In Section 4.3, we illustrate our results by showing how some renowned examples, including weighted random walks, bandit algorithms or decreasing step Euler schemes, can be easily studied with this unified approach. In Section 4.4 and 4.5, we provide the proofs of our main theorems and of the technical parts left aside while dealing with the illustrations.

4.2 Main results

4.2.1 Framework

We shall use the following notation in the sequel:

- $\mathscr{C}_b^N$ is the set of $\mathscr{C}^N(\mathbb{R}^D)$ functions such that $\sum_{j=0}^N \|f^{(j)}\|_\infty < +\infty$, for $N \in \mathbb{N} := \{0, 1, 2, \ldots\}$.
- $\mathscr{C}_c^N$ is the set of $\mathscr{C}^N(\mathbb{R}^D)$ functions with compact support, for $N \in \mathbb{N} \cup \{+\infty\}$.
- $\mathscr{C}_0^0 = \{f \in \mathscr{C}^0(\mathbb{R}^D) : \lim_{\|x\|\to\infty} f(x) = 0\}$.
- $\mathscr{L}(X)$ is the law of a random variable X and $\mathrm{Supp}(\mathscr{L}(X))$ its support.
- $x \wedge y := \min(x, y)$ and $x \vee y := \max(x, y)$ for any $x, y \in \mathbb{R}$.
- $f^{(j)}$ is the differential of order j of a function $f \in \mathscr{C}^j(\mathbb{R}^D)$, and
$$\|f^{(j)}\|_\infty = \sup_{|\alpha|=j} \sup_{x\in\mathbb{R}^D} |D^\alpha f(x)|.$$
- $\chi_d(x) := \sum_{k=0}^d \|x\|^k$ for $x \in \mathbb{R}^D$.

Let us recall some basics about Markov processes. Given a homogeneous Markov process $(X_t)_{t\geq 0}$ with right continuous with left limits (càdlàg) trajectories almost surely (a.s.), we define its Markov semigroup $(P_t)_{t\geq 0}$ by

$$P_t f(x) = \mathbb{E}[f(X_t) \mid X_0 = x].$$

It is said to be Feller if, for all $f \in \mathscr{C}_0^0$, $P_t f \in \mathscr{C}_0^0$ and $\lim_{t\to 0} \|P_t f - f\|_\infty = 0$. We can define its generator $\mathcal{L}$ acting on functions f satisfying $\lim_{t\to 0} \|t^{-1}(P_t f - f) - \mathcal{L} f\|_\infty = 0$. The set of such functions is denoted by $\mathcal{D}(\mathcal{L})$, and is dense in $\mathscr{C}_0^0$; see for instance [EK86]. The semigroup property of (P_t) ensures the existence of a semiflow

$$\Phi(\nu, t) := \nu P_t, \tag{4.2.1}$$

defined for any probability measure ν and $t \geq 0$; namely, for all $s, t > 0$, $\Phi(\nu, t + s) = \Phi(\Phi(\nu, t), s)$.

Let $(y_n)_{n\geq 0}$ be a (inhomogeneous) Markov chain and let $(\mathcal{L}_n)_{n\geq 0}$ be a sequence of operators satisfying, for $f \in \mathscr{C}_b^0$,

$$\mathcal{L}_n f(y_n) := \frac{\mathbb{E}\left[f(y_{n+1}) - f(y_n)|y_n\right]}{\gamma_{n+1}},$$

where $(\gamma_n)_{n\geq 1}$ is a decreasing sequence converging to 0, such that $\sum_{n=1}^{\infty} \gamma_n = +\infty$. Note that the sequence $(\mathcal{L}_n)$ exists thanks to Doob's lemma. Let (τ_n) be the sequence defined by $\tau_0 := 0$ and $\tau_n := \sum_{k=1}^{n} \gamma_k$, and let $m(t) := \sup\{n \geq 0 \ : \ t \geq \tau_n\}$ be the unique integer such that $\tau_{m(t)} \leq t < \tau_{m(t)+1}$. We denote by (Y_t) the process defined by $Y_t := y_n$ when $t \in [\tau_n, \tau_{n+1})$ and we set

$$\mu_t := \mathscr{L}(Y_t). \tag{4.2.2}$$

Following [BH96, Ben99], we say that $(\mu_t)_{t\geq 0}$ is an asymptotic pseudotrajectory of Φ (with respect to a distance d over probability distributions) if, for any $T > 0$,

$$\lim_{t\to\infty} \sup_{0\leq s\leq T} d(\mu_{t+s}, \Phi(\mu_t, s)) = 0. \tag{4.2.3}$$

Likewise, we say that $(\mu_t)_{t\geq 0}$ is a λ-pseudotrajectory of Φ (with respect to d) if there exists $\lambda > 0$ such that, for all $T > 0$,

$$\limsup_{t\to+\infty} \frac{1}{t} \log\left(\sup_{0\leq s\leq T} d(\mu_{t+s}, \Phi(\mu_t, s))\right) \leq -\lambda. \tag{4.2.4}$$

This definition of λ-pseudotrajectories is the same as in [Ben99], up to the sign of λ.

In the sequel, we discuss asymptotic pseudotrajectories with distances of the form

$$d_{\mathscr{F}}(\mu, \nu) := \sup_{f\in\mathscr{F}} |\mu(f) - \nu(f)| = \sup_{f\in\mathscr{F}} \left|\int f d\mu - \int f d\nu\right|,$$

for a certain class of functions $\mathscr{F}$. In particular, this includes total variation, Fortet-Mourier and Wasserstein distances. In general, $d_{\mathscr{F}}$ is a pseudodistance. Nevertheless, it is a distance whenever $\mathscr{F}$ contains an algebra of bounded continuous functions that separates points (see [EK86, Theorem 4.5.(a), Chapter 3]). In all the cases considered here, $\mathscr{F}$ contains the algebra $\mathscr{C}_c^{\infty}$ and then convergence in $d_{\mathscr{F}}$ entails convergence in distribution. Indeed, the following lemma holds (the proof is classical, and is given in the appendix in Section 4.5 for the sake of completeness).

Lemma 4.2.1 (*Weak convergence and $d_{\mathscr{F}}$*)

Assume that $\mathscr{F}$ is a star domain with respect to 0 (i.e. if $f \in \mathscr{F}$ then $\lambda f \in \mathscr{F}$ for $\lambda \in [0, 1]$). Let $(\mu_n), \mu$ be probability measures. If $\lim_{n\to\infty} d_{\mathscr{F}}(\mu_n, \mu) = 0$ and, for every $g \in \mathscr{C}_c^{\infty}$, there exists $\lambda > 0$ such that $\lambda g \in \mathscr{F}$, then (μ_n) converges weakly toward μ. If $\mathscr{F} \subseteq \mathscr{C}_b^1$, then $d_{\mathscr{F}}$ metrizes the weak convergence.

4.2.2 Assumptions and main theorem

In the sequel, let d_1, N_1, N_2 be non-negative integers, parameters of the model. We will assume, without loss of generality, that $N_1 \leq N_2$. Some key methods of how to check every assumption are provided in Section 4.3.

The first assumption we need is crucial. It defines the asymptotic homogeneous Markov process ruling the asymptotic behavior of (y_n).

Assumption 4.2.2 (*Convergence of generators*)

There exists a non-increasing sequence $(\epsilon_n)_{n\geq 1}$ converging to 0 and a constant M_1 (depending on $\mathscr{L}(y_0)$) such that, for all $f \in \mathcal{D}(\mathcal{L}) \cap \mathscr{C}_b^{N_1}$ and $n \in \mathbb{N}^\star$, and for any $y \in Supp(\mathscr{L}(y_n))$

$$|\mathcal{L}f(y) - \mathcal{L}_n f(y)| \leq M_1 \chi_{d_1}(y) \sum_{j=0}^{N_1} \|f^{(j)}\|_\infty \epsilon_n.$$

The following assumption is quite technical, but turns out to be true for most of the limit semigroups we deal with. Indeed, this is shown for large classes of PDMPs in Proposition 4.3.6 and for some diffusion processes in Lemma 4.3.12.

Assumption 4.2.3 (*Regularity of the limit semigroup*)

For all $T > 0$, there exists a constant C_T such that, for every $t \leq T, j \leq N_1$ and $f \in \mathscr{C}_b^{N_2}$,

$$P_t f \in \mathscr{C}_b^{N_1}, \quad |(P_t f)^{(j)}(y)| \leq C_T \sum_{i=0}^{N_2} \|f^{(i)}\|_\infty.$$

The next assumption is a standard condition of uniform boundedness of the moments of the Markov chain. We also provide a very similar Lyapunov criterion to check this condition.

Assumption 4.2.4 (*Uniform boundedness of moments*)

Assume that there exists an integer $d \geq d_1$ such that one of the following statements holds:

i) There exists a constant M_2 (depending on $\mathscr{L}(y_0)$) such that

$$\sup_{n\geq 0} \mathbb{E}[\chi_d(y_n)] \leq M_2.$$

ii) There exists $V : \mathbb{R}^D \to \mathbb{R}_+$ such that, for all $n \geq 0$, $\mathbb{E}[V(y_n)] < +\infty$. Moreover, there exist $n_0 \in \mathbb{N}^\star, a, \alpha, \beta > 0$, such that $V(y) \geq \chi_d(y)$ when $|y| > a$, such that, for $n \geq n_0$, and for any $y \in Supp(\mathscr{L}(y_n))$

$$\mathcal{L}_n V(y) \leq -\alpha V(y) + \beta.$$

In this assumption, the function V is a so-called Lyapunov function. The integer d can be thought of as $d = d_1$ (which is sufficient for Theorem 4.2.7 to hold). However, in the setting of Assumption 4.2.12, it might be necessary to consider $d > d_1$. Of course, if Assumption 4.2.4 holds for $d' > d$, then it holds for d. Note that we usually can take $V(y) = \mathrm{e}^{\theta y}$, so that we can choose d as large as needed.

Remark 4.2.5 (*ii)* ⇒ *i)*): Computing $\mathbb{E}[\chi_d(y_n)^d]$ to check Assumption 4.2.4.i) can be involved, so we rather check a Lyapunov criterion. It is classic that *ii)* entails *i)*. Indeed, denoting by $n_1 := n_0 \vee \min\{n \in \mathbb{N}^\star : \gamma_n < \alpha^{-1}\}$ and $v_n := \mathbb{E}[V(y_n)]$, it is clear that

$$v_{n+1} \leq v_n + \gamma_{n+1}(\beta - \alpha v_n).$$

From this inequality, it is easy to deduce that, for $n \geq n_1$, $v_{n+1} \leq \beta\alpha^{-1} \vee v_n$ and then by induction $v_n \leq \beta\alpha^{-1} \vee v_{n_1}$, which entails *i)*. Then,

$$\begin{aligned}\mathbb{E}[\chi_d(y_n)] &= \mathbb{P}(|y_n| \leq a)\mathbb{E}[\chi_d(y_n)||y_n| \leq a] + \mathbb{P}(|y_n| > a)\mathbb{E}[\chi_d(y_n)||y_n| > a] \\ &\leq \chi_d(a) + \frac{\beta}{\alpha} \vee \left(\sup_{k \leq n_1} v_k\right).\end{aligned}$$

♢

Note that, with a classical approach, Assumption 4.2.4 would provide tightness and Assumption 4.2.2 would be used to identify the limit.

The previous three assumptions are crucial to provide a result on asymptotic pseudotrajectories (Theorem 4.2.7), but are not enough to quantify speeds of convergence. As it can be observed in the proof of Theorem 4.2.7, such speed relies deeply on the asymptotic behavior of $\gamma_{m(t)}$ and $\epsilon_{m(t)}$. To this end, we follow the guidelines of [Ben99] to provide a condition in order to ensure such an exponential decay. For any non-increasing sequences $(\gamma_n), (\epsilon_n)$ converging to 0, define

$$\lambda(\gamma, \epsilon) = -\limsup_{n\to\infty} \frac{\log(\gamma_n \vee \epsilon_n)}{\sum_{k=1}^n \gamma_k}.$$

Remark 4.2.6 (*Computation of* $\lambda(\gamma, \epsilon)$): With the notation of [Ben99, Proposition 8.3], we have $\lambda(\gamma, \gamma) = -l(\gamma)$. It is easy to check that, if $\epsilon_n \leq \gamma_n$ for n large, $\lambda(\gamma, \epsilon) = \lambda(\gamma, \gamma)$ and, if $\epsilon_n = \gamma_n^\beta$ with $\beta \leq 1$, $\lambda(\gamma, \epsilon) = \beta\lambda(\gamma, \gamma)$. We can mimic [Ben99, Remark 8.4] to provide sufficient conditions for $\lambda(\gamma, \epsilon)$ to be positive. Indeed, if $\gamma_n = f(n), \epsilon_n = g(n)$ with f, g two positive functions decreasing toward 0 such that $\int_1^{+\infty} f(s)ds = +\infty$, then

$$\lambda(\gamma, \epsilon) = -\limsup_{x\to\infty} \frac{\log(f(x) \vee g(x))}{\int_1^x f(s)ds}.$$

Typically, if

$$\gamma_n \sim \frac{A}{n^a \log(n)^b}, \quad \epsilon_n \sim \frac{B}{n^c \log(n)^d}$$

for $A, B, a, b, c, d \geq 0$, then

- $\lambda(\gamma, \epsilon) = 0$ for $a < 1$.

- $\lambda(\gamma, \epsilon) = (c \wedge 1)A^{-1}$ for $a = 1$ and $b = 0$.
- $\lambda(\gamma, \epsilon) = +\infty$ for $a = 1$ and $0 < b \leq 1$.

$\diamondsuit$

Now, let us provide the main results of this paper.

Theorem 4.2.7 (*Asymptotic pseudotrajectories*)

Let $(y_n)_{n\geq 0}$ be an inhomogeneous Markov chain and let Φ and μ be defined as in (4.2.1) and (4.2.2). If Assumptions 4.2.2, 4.2.3, 4.2.4 hold, then $(\mu_t)_{t\geq 0}$ is an asymptotic pseudotrajectory of Φ with respect to $d_{\mathscr{F}}$, where

$$\mathscr{F} = \left\{ f \in \mathcal{D}(\mathcal{L}) \cap \mathscr{C}_b^{N_2} : \mathcal{L}f \in \mathcal{D}(\mathcal{L}), \|\mathcal{L}f\|_\infty + \|\mathcal{L}\mathcal{L}f\|_\infty + \sum_{j=0}^{N_2} \|f^{(j)}\|_\infty \leq 1 \right\}.$$

Moreover, if $\lambda(\gamma, \epsilon) > 0$, then $(\mu_t)_{t\geq 0}$ is a $\lambda(\gamma, \epsilon)$-pseudotrajectory of Φ with respect to $d_{\mathscr{F}}$.

4.2.3 Consequences

Theorem 4.2.7 relates the asymptotic behavior of the Markov chain (y_n) to the one of the Markov process generated by $\mathcal{L}$. However, to deduce convergence or speeds of convergence of the Markov chain, we need another assumption:

Assumption 4.2.8 (*Ergodicity*)

Assume that there exist a probability distribution π, constants $v, M_3 > 0$ (M_3 depending on $\mathscr{L}(y_0)$), and a class of functions $\mathscr{G}$ such that one of the following conditions holds:

i) $\mathscr{G} \subseteq \mathscr{F}$ *and, for any probability measure ν, for all $t > 0$,*
$$d_{\mathscr{G}}(\Phi(\nu, t), \pi) \leq d_{\mathscr{G}}(\nu, \pi) M_3 e^{-vt}.$$

ii) There exists $r, M_4 > 0$ such that, for all $s, t > 0$
$$d_{\mathscr{G}}(\Phi(\mu_s, t), \pi) \leq M_3 e^{-vt},$$
and, for all $T > 0$, with C_T defined in Assumption 4.2.3,
$$TC_T \leq M_4 e^{rT}.$$

iii) There exist functions $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ and $W \in \mathscr{C}^0$ such that
$$\lim_{t\to\infty} \psi(t) = 0, \qquad \lim_{\|x\|\to\infty} W(x) = +\infty, \qquad \sup_{n\geq 0} \mathbb{E}[W(y_n)] < \infty,$$
and, for any probability measure ν, for all $t \geq 0$,
$$d_{\mathscr{G}}(\Phi(\nu, t), \pi) \leq \nu(W)\psi(t).$$

Since standard proofs of geometric ergodicity rely on the use of Grönwall's Lemma, Assumption 4.2.8.i) and ii) are quite classic. In particular, using Foster-Lyapunov methods entails such inequalities (see e.g. [MT93b, HM11]). However, in a weaker setting (sub-geometric ergodicity for instance) Assumption 4.2.8.iii) might still hold; see for example [Hai10, Theorem 4.1] or [DFG09, Theorem 3.10]. Note that, if $W = \chi_d$, then $\sup_{n\geq 0} \mathbb{E}[W(y_n)] < \infty$ automatically from Assumption 4.2.4. Note that, in classical settings where $TC_T \leq M_4 e^{rT}$, we have *i)*$\Rightarrow$ *ii)*$\Rightarrow$ *iii)*.

Theorem 4.2.9 (*Speed of convergence toward equilibrium*)

Assume that Assumptions 4.2.2, 4.2.3, 4.2.4 hold and let $\mathscr{F}$ be as in Theorem 4.2.7.

i) If Assumption 4.2.8.i) holds and $\lambda(\gamma,\epsilon) > 0$ then, for any $u < \lambda(\gamma,\epsilon) \wedge v$, there exists a constant M_5 such that, for all $t > t_0 := (v-u)^{-1}\log(1 \wedge M_3)$,

$$d_{\mathscr{G}}(\mu_t, \pi) \leq (M_5 + d_{\mathscr{G}}(\mu_0, \pi))\, e^{-ut}.$$

ii) If Assumption 4.2.8.ii) holds and $\lambda(\gamma,\epsilon) > 0$ then, for any $u < v\lambda(\gamma,\epsilon)(r + v + \lambda(\gamma,\epsilon))^{-1}$, there exists a constant M_5 such that, for all $t > 0$,

$$d_{\mathscr{F}\cap\mathscr{G}}(\mu_t, \pi) \leq M_5 e^{-ut}.$$

iii) If Assumption 4.2.8.iii) holds and convergence in $d_{\mathscr{G}}$ implies weak convergence, then μ_t converges weakly toward π when $t \to \infty$.

The first part of this theorem is similar to [Ben99, Lemma 8.7] but provides sharp bounds for the constants. In particular, M_5 and t_0 do not depend on μ_0 (in Theorem 4.2.9.*i)* only), see the proof for an explicit expression of M_5). The second part, however, does not require $\mathscr{G}$ to be a subset of $\mathscr{F}$, which can be rather involved to check, given the expression of $\mathscr{F}$ given in Theorem 4.2.7. The third part is a direct consequence of [Ben99, Theorem 6.10].

Remark 4.2.10 (*Rate of convergence in the initial scale*): Theorem 4.2.9.i) and ii) provide a bound of the form

$$d_{\mathscr{H}}(\mathcal{L}(Y_t), \pi) \leq Ce^{-ut},$$

for some $\mathscr{H}, C, u$ and all $t \geq 0$. This easily entails, for another constant C and all $n \geq 0$,

$$d_{\mathscr{H}}(\mathcal{L}(y_n), \pi) \leq Ce^{-u\tau_n}.$$

Let us detail this bound for three examples where $\epsilon \leq \gamma$:

- if $\gamma_n = An^{-1/2}$, then $d_{\mathscr{H}}(\mathcal{L}(y_n), \pi) \leq Ce^{-2Au\sqrt{n}}$.
- if $\gamma_n = An^{-1}$, then $d_{\mathscr{H}}(\mathcal{L}(y_n), \pi) \leq Cn^{-Au}$.
- if $\gamma_n = A(n\log(n))^{-1}$, then $d_{\mathscr{H}}(\mathcal{L}(y_n), \pi) \leq C\log(n)^{-Au}$.

In a nutshell, if γ_n is large, the speed of convergence is good but $\lambda(\gamma,\gamma)$ is small. In particular, even if $\gamma_n = n^{-1/2}$ provides the better speed, Theorem 4.2.9 does not apply. Remark that the parameter u is more important at the discrete time scale than it is at the continuous time scale. $\diamondsuit$

Remark 4.2.11 (*Convergence of unbounded functionals*): Theorem 4.2.9 provides convergence in distribution of (μ_t) toward π, i.e. for every $f \in \mathscr{C}_b^0(\mathbb{R}^D)$,

$$\lim_{t\to\infty} \mu_t(f) = \pi(f).$$

Nonetheless, Assumption 4.2.4 enables us to extend this convergence to unbounded functionals f. Recall that, if a sequence $(X_n)_{n\geq 0}$ converges weakly to X and

$$M := \mathbb{E}[V(X)] + \sup_{n\geq 0} \mathbb{E}[V(X_n)] < +\infty$$

for some positive function V, then $\mathbb{E}[f(X_n]$ converges to $\mathbb{E}[f(X)]$ for every function $|f| < V^\theta$, with $\theta < 1$. Indeed, let $(\varphi_k)_{k\geq 0}$ be a sequence of $\mathscr{C}_c^\infty$ functions such that $\forall x \in \mathbb{R}^D, \lim_{k\to\infty} \varphi_k(x) = 1$ and $0 \leq \varphi_k \leq 1$. We have, for $k \in \mathbb{N}$,

$$\begin{aligned}
|\mathbb{E}\left[f(X_n) - f(X)\right]| &\leq |\mathbb{E}\left[(1-\varphi_k(X_n))f(X_n)\right]| + |\mathbb{E}\left[(1-\varphi_k(X))f(X)\right]| \\
&\quad + |\mathbb{E}\left[f(X_n)\varphi_k(X_n) - f(X)\varphi_k(X)\right]| \\
&\leq \mathbb{E}[|f(X_n)|^{\frac{1}{\theta}}]^\theta \mathbb{E}[(1-\varphi_k(X_n))^{\frac{1}{1-\theta}}]^{1-\theta} \\
&\quad + \mathbb{E}[|f(X)|^{\frac{1}{\theta}}]^\theta \mathbb{E}[(1-\varphi_k(X))^{\frac{1}{1-\theta}}]^{1-\theta} \\
&\quad + |\mathbb{E}\left[f(X_n)\varphi_k(X_n) - f(X)\varphi_k(X)\right]| \\
&\leq M^\theta \mathbb{E}[(1-\varphi_k(X_n))^{\frac{1}{1-\theta}}]^{1-\theta} + M^\theta \mathbb{E}[(1-\varphi_k(X))^{\frac{1}{1-\theta}}]^{1-\theta} \\
&\quad + |\mathbb{E}\left[f(X_n)\varphi_k(X_n) - f(X)\varphi_k(X)\right]|,
\end{aligned}$$

so that, for all $k \in \mathbb{N}$,

$$\limsup_{n\to\infty} \mathbb{E}\left[f(X_n) - f(X)\right] \leq 2M^\theta \mathbb{E}[(1-\varphi_k(X))^{\frac{1}{1-\theta}}]^{1-\theta}.$$

Using the dominated convergence theorem, $\lim_{n\to\infty} \mathbb{E}\left[f(X_n) - f(X)\right] = 0$ since the right-hand side converges to 0. Note that the condition $|f| \leq V^\theta$ can be slightly weakened using the generalized Hölder's inequality on Orlicz spaces (see e.g. [CGLP12]). Although, note that $\mathbb{E}[V(X_n)]$ may not converge to $\mathbb{E}[V(X)]$. $\diamondsuit$

The following assumption is purely technical but is easy to verify in all of our examples, and will be used to prove functional convergence.

Assumption 4.2.12 (*Control of the variance*)

Define the following operator:

$$\Gamma_n f = \mathcal{L}_n f^2 - \gamma_{n+1}(\mathcal{L}_n f)^2 - 2f\mathcal{L}_n f.$$

Assume that there exists $d_2 \in \mathbb{N}$ and $M_6 > 0$ such that, if φ_i is the projection on the i^{th} coordinate,

$$\mathcal{L}_n \varphi_i(y) \leq M_6 \chi_{d_2}(y), \quad \Gamma_n \varphi_i(y) \leq M_6 \chi_{d_2}(y),$$

and

$$\mathcal{L}_n \chi_{d_2}(y) \leq M_6 \chi_{d_2}(y), \quad \Gamma_n \chi_{d_2}(y) \leq M_6 \chi_d(y).$$

Theorem 4.2.13 (*Functional convergence*)

Assume that Assumptions 4.2.2, 4.2.3, 4.2.4, 4.2.8 hold and let π be as in Assumption 4.2.8. Let $Y_s^{(t)} := Y_{t+s}$ and X^π be the process generated by $\mathcal{L}$ such that $\mathscr{L}(X_0^\pi) = \pi$. Then, for any $m \in \mathbb{N}^\star$, let $0 < s_1 < \cdots < s_m$,

$$(Y_{s_1}^{(t)}, \ldots, Y_{s_m}^{(t)}) \xrightarrow{\mathscr{L}} (X_{s_1}^\pi, \ldots, X_{s_m}^\pi).$$

Moreover, if Assumption 4.2.12 holds, then the sequence of processes $(Y_s^{(t)})_{s\geq 0}$ converges in distribution, as $t \to +\infty$, toward $(X_s^\pi)_{s\geq 0}$ in the Skorokhod space.

For reminders about the Skorokhod space, the reader may consult [JM86, Bil99, JS03]. Note that the operator Γ_n we introduced in Assumption 4.2.12 is very similar to the *carré du champ* operator in the continuous-time case, up to a term $\gamma_{n+1}(\mathcal{L}_n f)^2$ vanishing as $n \to +\infty$ (see e.g. [Bak94, ABC+00, JS03]). Moreover, if we denote by (K_n) the transition kernels of the Markov chain (y_n), then it is clear that

$$\forall n \in \mathbb{N}, \quad \gamma_{n+1}\Gamma_n f = K_n f^2 - (K_n f)^2.$$

4.3 Illustrations

4.3.1 Weighted Random Walks

In this section, we apply Theorems 4.2.7 and 4.2.9 to the *Weighted Random Walk* (WRW) on $\mathbb{R}^D$. Let (ω_n) be a positive sequence, and $\gamma_n := \omega_n(\sum_{k=1}^n \omega_k)^{-1}$. Then, set

$$x_n := \frac{\sum_{k=1}^n \omega_k E_k}{\sum_{k=1}^n \omega_k}, \quad x_{n+1} := x_n + \gamma_{n+1}\left(E_{n+1} - x_n\right).$$

Here, x_n is the weighted mean of $E_1, \ldots, E_n$, where (E_n) is a sequence of centered independent random variables. Under standard assumptions on the moments of E_n, the strong law of large numbers holds and (x_n) converges to 0 a.s. Thus, it is natural to apply the general setting of Section 4.2 to $y_n := x_n \gamma_n^{-1/2}$ and to define μ_t as in (4.2.2). As we shall see, computations lead to the convergence of $\mathcal{L}_n$, as defined in (4.1.1), toward

$$\mathcal{L}f(y) := -yl(\gamma)f'(y) + \frac{\sigma^2}{2} f''(y),$$

where $l(\gamma)$ and σ are defined below. Hence, the properly normalized process asymptotically behaves like the Ornstein-Uhlenbeck process; see Figure 4.3.1. This process is the solution of the following Stochastic Differential Equation (SDE):

$$dX_t = -l(\gamma)X_t dt + \sigma dW_t,$$

see [Bak94] for instance. In the sequel, define $\mathscr{F}$ as in Theorem 4.2.7 with $N_2 = 3$, and φ_i the projection on the i^{th} coordinate.

Proposition 4.3.1 (*Results for the WRW*)

Assume that

$$\mathbb{E}\left[\sum_{i=1}^{D}\varphi_i(E_{n+1})^2\right]=\sigma^2, \quad \sup_{n\geq 1}\gamma_n^2\omega_n^4\mathbb{E}[\|E_n\|^4]<+\infty, \quad \sup_{n}\gamma_n\sum_{i=1}^{n}\omega_i^2<+\infty,$$

and that there exist $l(\gamma)>0$ and $\beta(\gamma)>1$ such that

$$\sqrt{\frac{\gamma_n}{\gamma_{n+1}}}-1-\sqrt{\gamma_n\gamma_{n+1}}=-\gamma_n l(\gamma)+\mathcal{O}(\gamma_n^{\beta(\gamma)}). \tag{4.3.1}$$

Then (μ_t) is an asymptotic pseudotrajectory of Φ, with respect to $d_{\mathscr{F}}$.

Moreover, if $\lambda(\gamma,\gamma^{(\beta(\gamma)-1)\wedge\frac{1}{2}})>0$ then, for any $u<l(\gamma)\lambda(\gamma,\gamma^{(\beta(\gamma)-1)\wedge\frac{1}{2}})(l(\gamma)+\lambda(\gamma,\gamma^{(\beta(\gamma)-1)\wedge\frac{1}{2}}))^{-1}$, there exists a constant C such that, for all $t>0$,

$$d_{\mathscr{F}}(\mu_t,\pi)\leq Ce^{-ut}, \tag{4.3.2}$$

where π is the Gaussian distribution $\mathscr{N}(0,\sigma^2/(2l(\gamma)))$.

Moreover, the sequence of processes $(Y_s^{(t)})_{s\geq 0}$ converges in distribution, as $t\to+\infty$, toward $(X_s^\pi)_{s\geq 0}$ in the Skorokhod space.

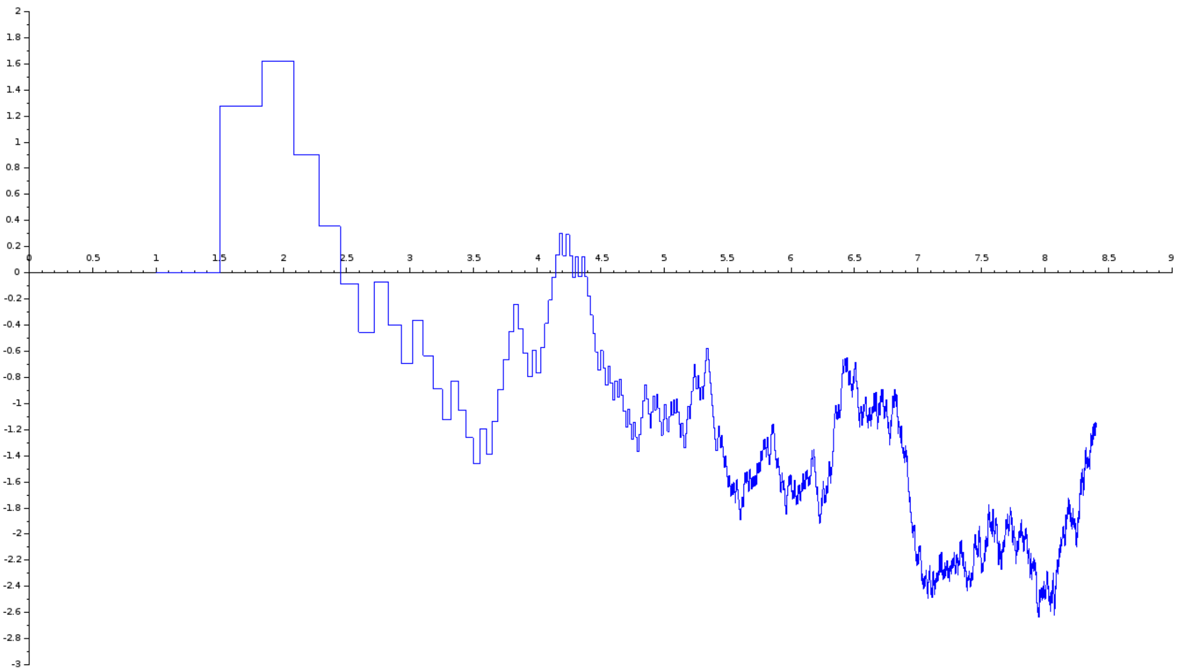


Figure 4.3.1 – Trajectory of the interpolated process for the normalized mean of the WRW with $\omega_n=1$ and $\mathscr{L}(E_n)=(\delta_{-1}+\delta_1)/2$.

It is possible to recover the functional convergence using classical results: for instance, one can apply [KY03, Theorem 2.1, Chapter 10] with a slightly stronger assumption on (γ_n). Yet, to our knowledge, the rate of convergence (4.3.2) is original.

Remark 4.3.2 (*Powers of* n): Typically, if $\gamma_n\sim An^{-\alpha}$, then we can easily check that

- if $\alpha = 1$, then (4.3.1) holds with $l(\gamma) = 1 - \frac{1}{2A}$ and $\beta(\gamma) = 2$.
- if $0 < \alpha < 1$, then (4.3.1) holds with $l(\gamma) = 1$ and $\beta(\gamma) = \frac{1+\alpha}{\alpha} > 2$.

Observe that, if $\omega_n = n^a$ for any $a > -1$, then $\gamma_n \sim \frac{1+a}{n}$ and (4.3.1) holds with $l(\gamma) = \frac{1+2a}{2+2a}$ and $\beta(\gamma) = 2$. ♢

We will see during the proof that checking Assumptions 4.2.2, 4.2.3, 4.2.4 and 4.2.8 is quite direct.

Proof of Proposition 4.3.1: For the sake of simplicity, we do the computations for $D = 1$. We have

$$y_{n+1} = \sqrt{\frac{\gamma_n}{\gamma_{n+1}}} y_n + \sqrt{\gamma_{n+1}}(E_{n+1} - \sqrt{\gamma_n} y_n),$$

so

$$\mathcal{L}_n f(y) = \gamma_{n+1}^{-1} \mathbb{E}[f(y_{n+1}) - f(y_n) | y_n = y] = \gamma_{n+1}^{-1} \mathbb{E}[f(y + I_n(y)) - f(y)],$$

with $I_n(y) := \left(\sqrt{\frac{\gamma_n}{\gamma_{n+1}}} - 1 - \sqrt{\gamma_n \gamma_{n+1}}\right) y + \sqrt{\gamma_{n+1}} E_{n+1}$. Simple Taylor expansions provide the following equalities (where $\mathcal{O}$ is the Landau notation, deterministic and uniform over y and f, and $\beta := \beta(\gamma) \wedge \frac{3}{2}$):

$$\begin{aligned}
I_n(y) &= \left(-\gamma_n l(\gamma) + \mathcal{O}(\gamma_n^\beta)\right) y + \sqrt{\gamma_{n+1}} E_{n+1}, \\
I_n^2(y) &= \gamma_{n+1} E_{n+1}^2 + \chi_2(y)(1 + E_{n+1}) \mathcal{O}\left(\gamma_{n+1}^\beta\right), \\
I_n^3(y) &= \chi_3(y)(1 + E_{n+1} + E_{n+1}^2 + E_{n+1}^3) \mathcal{O}\left(\gamma_{n+1}^\beta\right).
\end{aligned}$$

In the setting of Remark 4.3.2, note that $\beta = \frac{3}{2}$. Now, Taylor formula provides a random variable ξ_n^y such that

$$f(y + I_n(y)) - f(y) = I_n(y) f'(y) + \frac{I_n^2(y)}{2} f''(y) + \frac{I_n^3(y)}{6} f^{(3)}(\xi_n^y).$$

Then, it follows that

$$\begin{aligned}
\mathcal{L}_n f(y) &= \gamma_{n+1}^{-1} \mathbb{E}\left[I_n(y) f'(y) + \frac{I_n^2(y)}{2} f''(y) + \frac{I_n^3(y)}{6} f^{(3)}(\xi_n^y) \middle| y_n = y \right] \\
&= \gamma_{n+1}^{-1} \left[\left(-\gamma_n l(\gamma) + \mathcal{O}(\gamma_n^{3/2})\right) y + \sqrt{\gamma_{n+1}} \mathbb{E}[E_{n+1}] \right] f'(y) \\
&\quad + \frac{1}{2\gamma_{n+1}} \left[\gamma_{n+1} \mathbb{E}[E_{n+1}^2 + \chi_2(y) \mathcal{O}\left(\gamma_{n+1}^\beta\right) \right] f''(y) \\
&\quad + \gamma_{n+1}^{-1} \chi_3(y) \mathbb{E}[1 + E_{n+1} + E_{n+1}^2 + E_{n+1}^3] \| f^{(3)} \|_\infty \mathcal{O}\left(\gamma_{n+1}^\beta\right) \\
&= -y l(\gamma) f'(y) + \chi_1(y) \| f' \|_\infty \mathcal{O}\left(\gamma_n^{\beta-1}\right) + \frac{\sigma^2}{2} f''(y) + \chi_2(y) \| f'' \|_\infty \mathcal{O}\left(\gamma_n^{\beta-1}\right) \\
&\quad + \chi_3(y) \| f^{(3)} \|_\infty \mathcal{O}\left(\gamma_n^{\beta-1}\right). \qquad (4.3.3)
\end{aligned}$$

From (4.3.3), we can conclude that

$$|\mathcal{L}_n f(y) - \mathcal{L} f(y)| = \chi_3(y)(\| f' \|_\infty + \| f'' \|_\infty + \| f^{(3)} \|_\infty) \mathcal{O}(\gamma_n^{\beta-1}).$$

As a consequence, the WRW satisfies Assumptions 4.2.2 with $d_1 = 3$, $N_1 = 3$ and $\epsilon_n = \gamma_n^{\beta-1}$. Note that (see Remark 4.2.6) $\lambda(\gamma, \epsilon) = \beta(\gamma) - 1$ if $\gamma_n = n^{-1}$.

Now, let us show that $P_t f$ admits bounded derivatives for $f \in \mathscr{F}$. Here, the expressions of the semigroup and its derivatives are explicit and the computations are simple (see [Bak94, ABC+00]). Indeed, $P_t f(x) = \mathbb{E}[f(x\mathrm{e}^{-l(\gamma)t} + \sqrt{1-\mathrm{e}^{-2l(\gamma)t}}G)]$ and $(P_t f)^{(j)}(y) = \mathrm{e}^{-jl(\gamma)t} P_t f^{(j)}(y)$, where $\mathscr{L}(G) = \mathscr{N}(0,1)$. Then, it is clear that

$$\|(P_t f)^{(j)}\|_\infty = \mathrm{e}^{-jl(\gamma)t} \|P_t f^{(j)}\|_\infty \leq \|f^{(j)}\|_\infty.$$

Hence Assumption 4.2.3 holds with $N_2 = 3$ and $C_T = 1$. Without loss of generality (in order to use Theorem 4.2.13 later) we set $d = 4$.

Now, we check that the moments of order 4 of y_n are uniformly bounded. Applying Cauchy-Schwarz's inequality:

$$\mathbb{E}\left[\left\|\sum_{i=1}^n \omega_i E_i\right\|^4\right] = \mathbb{E}\left[\sum_{i=1}^n \omega_i^4 \|E_i\|^4 + 6\sum_{i<j} \omega_i^2 \|E_i^2\| \omega_j^2 \|E_j\|^2\right] \leq C\left(\sum_{i=1}^n \omega_i^2\right)^2,$$

for some explicit constant C. Then, since

$$\mathbb{E}[\|y_n\|^4] = \gamma_n^2 \mathbb{E}\left[\left\|\sum_{i=1}^n \omega_i E_i\right\|^4\right] \leq C \sup_{n\geq 1}\left(\gamma_n \sum_{i=1}^n \omega_i^2\right)^2,$$

the sequence $(y_n)_{n\geq 0}$ satisfies Assumption 4.2.4.

It is classic, using coupling methods with the same Brownian motion for instance, that, for any probability measure ν,

$$d_{\mathscr{G}}(\Phi(\nu,t),\pi) \leq d_{\mathscr{G}}(\nu,\pi)\mathrm{e}^{-l(\gamma)t},$$

where $\pi = \mathscr{N}(0, \sigma^2/(2l(\gamma))I_D)$ and $d_{\mathscr{G}}$ is the Wasserstein distance ($\mathscr{G}$ is the set of 1-Lipschitz functions, see [Che04]). We have, for $s,t > 0$,

$$d_{\mathscr{G}}(\Phi(\mu_s,t),\pi) \leq d_{\mathscr{G}}(\mu_s,\pi)\mathrm{e}^{-l(\gamma)t} \leq (M_2 + \pi(\chi_1))\mathrm{e}^{-l(\gamma)t}.$$

In other words, Assumption 4.2.8.ii) holds for the WRW model with $M_3 = M_2 + \pi(\chi_1), M_4 = 1, v = l(\gamma), r = 0$ and $\mathscr{F} \subseteq \mathscr{G}$.

Finally, it is easy to check Assumption 4.2.12 in the case of the WRW, with $d_2 = 2$, and then $\Gamma_n \chi_2 \leq M_6 \chi_4$ (that is why we set $d = 4$ above).

Then, Theorems 4.2.7, 4.2.9 and 4.2.13 achieve the proof of Proposition 4.3.1. □

Remark 4.3.3 (*Building a limit process with jumps*): In this paper, we mainly provide examples of Markov chains converging (in the sense of Theorem 4.2.7) toward diffusion processes (see Section 4.3.1) or jump processes (see Section 4.3.2). However, it is not hard to adapt the previous model to obtain an exemple converging toward a diffusion process with jumps (see Figure 4.3.2): this illustrates how every component (drift, jump and noise) appears in the limit generator. The intuition is that the jump terms appear when larger and larger jumps of the Markov chain occur with smaller and smaller probability. For an example when $D = 1$, take

$$\omega_n := 1, \quad E_n := \begin{cases} F_n & \text{if } U_n \geq \sqrt{\gamma_n} \\ \gamma_n^{-1/2} G_n & \text{if } U_n < \sqrt{\gamma_n} \end{cases}, \quad y_n := \frac{1}{\sqrt{\gamma_n}} \sum_{k=1}^{n} E_k,$$

where $(F_n)_{n\geq 1}, (G_n)_{n\geq 1}$ and $(U_n)_{n\geq 1}$ are three sequences of independent and identically distributed (i.i.d.) random variables, such that $\mathbb{E}[F_1] = 0, \mathbb{E}[F_1^2] = \sigma^2, \mathscr{L}(G_1) = Q$, $\mathscr{L}(U_1)$ is the uniform distribution on $[0, 1]$. In this case, $\gamma_n = 1/n$ and it is easy to show that $\mathcal{L}_n$ as defined in (4.1.1) converges toward the following infinitesimal generator:

$$\mathcal{L}f(y) := -\frac{1}{2} y f'(y) + \frac{\sigma^2}{2} f''(y) + \int_{\mathbb{R}} [f(y+z) - f(y)] Q(dz),$$

so that Assumption 4.2.2 holds with $d_1 = 3$, $N_1 = 3, \epsilon_n = n^{-1/2}$.

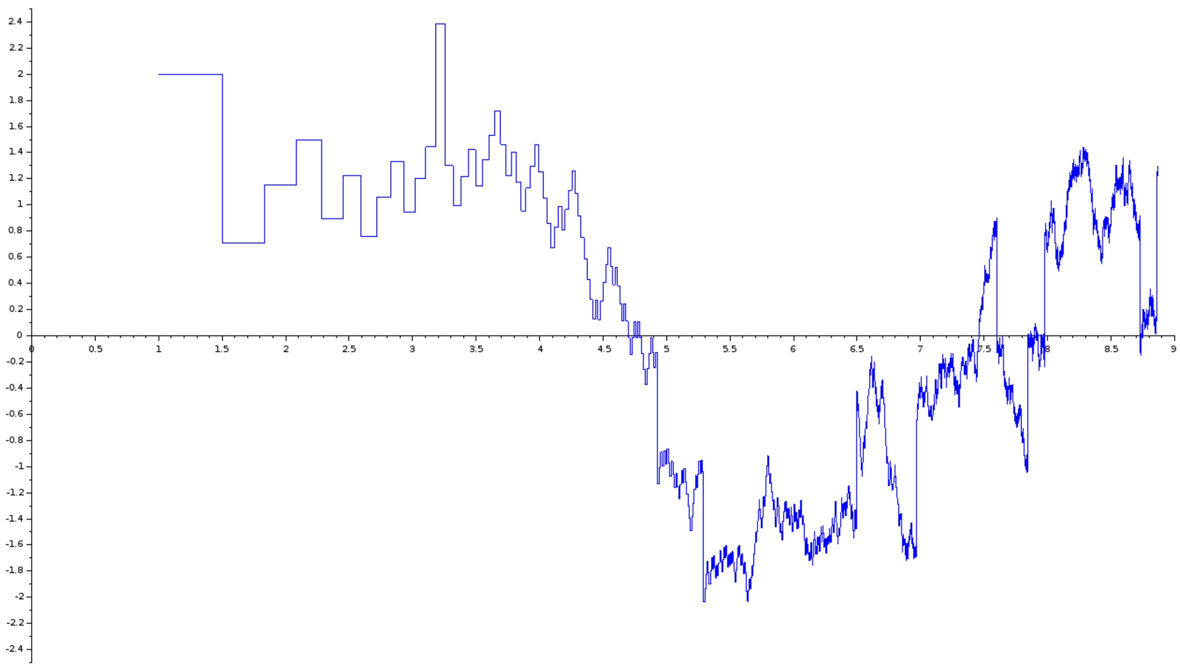


Figure 4.3.2 – Trajectory of the interpolated process for the toy model of Remark 4.3.3 with $\mathscr{L}(F_n) = \mathscr{L}(G_n) = (\delta_{-1} + \delta_1)/2$.

$\diamondsuit$

4.3.2 Penalized Bandit Algorithm

In this section, we slightly generalize the *Penalized Bandit Algorithm* (PBA) model introduced by Lamberton and Pagès, and we recover [LP08b, Theorem 4]. Such algorithms aim at optimizing the gain in a game with two choices, A and B, with respective unknown gain probabilities p_A and p_B. Originally, A and B are the two arms of a slot machine, or *bandit*. Throughout this section, we assume $0 \leq p_B < p_A \leq 1$.

Let $s : [0,1] \to [0,1]$ be a function, which can be understood as a player's strategy, such that $s(0) = 0, s(1) = 1$. Let $x_n \in [0,1]$ be a measure of her trust level in A at time n. She chooses A with probability $s(x_n)$ independently from the past, and updates x_n as follows:

x_{n+1}	Choice	Result
$x_n + \gamma_{n+1}(1 - x_n)$	A	Gain
$x_n - \gamma_{n+1}x_n$	B	Gain
$x_n + \gamma_{n+1}^2(1 - x_n)$	B	Loss
$x_n - \gamma_{n+1}^2 x_n$	A	Loss

Then (x_n) satisfies the following Stochastic Approximation algorithm:

$$x_{n+1} := x_n + \gamma_{n+1}\left(X_{n+1} - x_n\right) + \gamma_{n+1}^2\left(\widetilde{X}_{n+1} - x_n\right),$$

where

$$(X_{n+1}, \widetilde{X}_{n+1}) := \begin{cases} (1, x_n) & \text{with probability } p_1(x_n) \\ (0, x_n) & \text{with probability } p_0(x_n) \\ (x_n, 1) & \text{with probability } \widetilde{p}_1(x_n) \\ (x_n, 0) & \text{with probability } \widetilde{p}_0(x_n) \end{cases}, \tag{4.3.4}$$

with

$$p_1(x) = s(x)p_A, \quad p_0(x) = (1-s(x))p_B, \quad \widetilde{p}_1(x) = (1-s(x))(1-p_B), \quad \widetilde{p}_0(x) = s(x)(1-p_A). \tag{4.3.5}$$

Note that the PBA of [LP08b] is recovered by setting $s(x) = x$ in (4.3.5).

From now on, we consider the algorithm (4.3.4) where $p_1, p_0, \widetilde{p}_1, \widetilde{p}_0$ are non-necessarily given by (4.3.5), but are general non-negative functions whose sum is 1. Let $\mathscr{F}$ be as in Theorem 4.2.7 with $N_2 = 2$, and $y_n := \gamma_n^{-1}(1 - x_n)$ the rescaled algorithm. Let $\mathcal{L}_n$ be defined as in (4.1.1),

$$\mathcal{L}f(y) := [\widetilde{p}_0(1) - yp_1(1)]f'(y) - yp_0'(1)[f(y+1) - f(y)], \tag{4.3.6}$$

and π the invariant distribution for $\mathcal{L}$ (which exists and is unique, see Remark 4.3.7).

Under the assumptions of Proposition 4.3.4, it is straightforward to mimic the results [LP08b] and ensure that our generalized algorithm $(x_n)_{n\geq 0}$ satisfies the Ordinary Differential Equation (ODE) Lemma (see e.g. [KY03, Theorem 2.1, Chapter 5]), and converges toward 1 almost surely.

Proposition 4.3.4 (***Results for the PBA***)

Assume that $\gamma_n = n^{-1/2}$, that $p_1, \widetilde{p}_1, \widetilde{p}_0 \in C_b^1, p_0 \in C_b^2$, and that

$$p_0(1) = \widetilde{p}_1(1) = 0, \quad p_0'(1) \leq 0, \quad p_1(1) + p_0'(1) > 0, \quad \widetilde{p}_1(0) > 0.$$

If, for $0 < x < 1$, $(1-x)p_1(x) > xp_0(x)$, then (μ_t) is an asymptotic pseudotrajectory of Φ, with respect to $d_{\mathscr{F}}$.

Moreover, (μ_t) converges to π and the sequence of processes $(Y_s^{(t)})_{s\geq 0}$ converges in distribution, as $t \to +\infty$, toward $(X_s^\pi)_{s\geq 0}$ in the Skorokhod space.

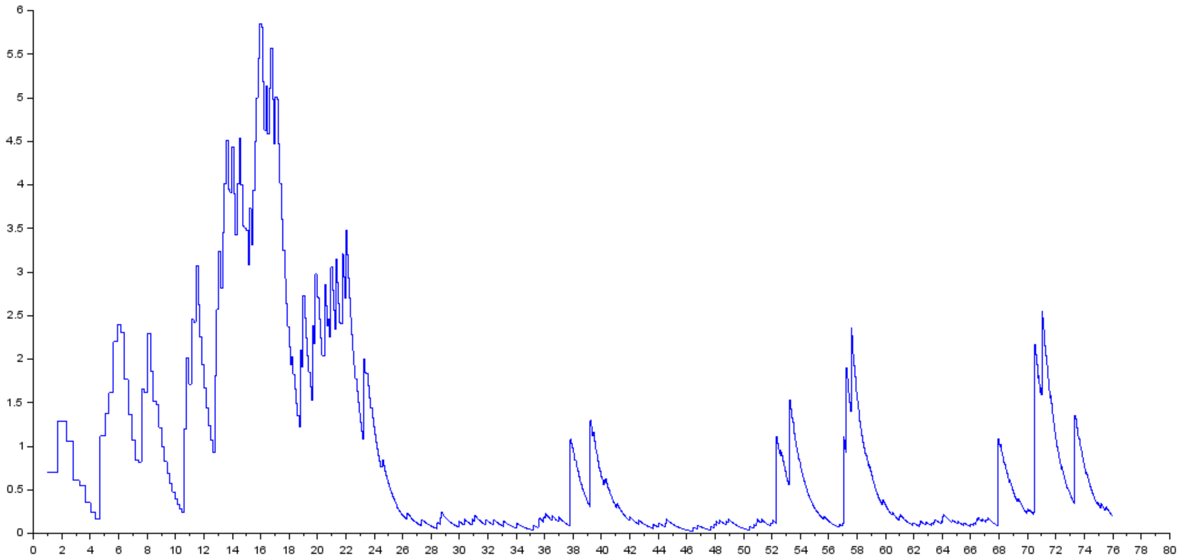


Figure 4.3.3 – Trajectory of the interpolated process for the rescaled PBA, setting $s(x) = x$ in (4.3.5).

The proof is given at the end of the section; before that, let us give some interpretation and heuristic explanation of the algorithm. The random sequence (y_n) satisfies

$$y_{n+1} = y_n + \left(\frac{\gamma_n}{\gamma_{n+1}} - 1\right) y_n - (X_{n+1} - x_n) - \gamma_{n+1}(\widetilde{X}_{n+1} - x_n),$$

thus, defining $\mathcal{L}_n$ as in (4.1.1),

$$\mathcal{L}_n f(y) = \gamma_{n+1}^{-1}\mathbb{E}\left[f(y_{n+1}) - f(y_n)\right| y_n = y] = \gamma_{n+1}^{-1}\mathbb{E}\left[f(y + I_n(y)) - f(y)|y_n = y\right],$$

where

$$I_n(y) := \begin{cases} I_n^1(y) := \left(\frac{\gamma_n}{\gamma_{n+1}} - 1 - \gamma_n\right) y & \text{with probability } p_1(1-\gamma_n y) \\ I_n^0(y) := 1 + \left(\frac{\gamma_n}{\gamma_{n+1}} - 1 - \gamma_n\right) y & \text{with probability } p_0(1-\gamma_n y) \\ \widetilde{I}_n^1(y) := \left(\frac{\gamma_n}{\gamma_{n+1}} - 1 - \gamma_n\gamma_{n+1}\right) y & \text{with probability } \widetilde{p}_1(1-\gamma_n y) \\ \widetilde{I}_n^0(y) := \gamma_{n+1} + \left(\frac{\gamma_n}{\gamma_{n+1}} - 1 - \gamma_n\gamma_{n+1}\right) y & \text{with probability } \widetilde{p}_0(1-\gamma_n y) \end{cases}. \tag{4.3.7}$$

Taylor expansions provide the convergence of $\mathcal{L}_n$ toward $\mathcal{L}$. As a consequence, the properly renormalized interpolated process will asymptotically behave like a PDMP (see Figure 4.3.3). Classically, one can read the dynamics of the limit process through its generator (see e.g. [Dav93]): the PDMP generated by (4.3.6) has upward jumps of height 1 and follows the flow given by the ODE $y' = \widetilde{p}_0(1) - yp_1(1)$, which means it converges exponentially fast toward $\widetilde{p}_0(1)/p_1(1)$.

Remark 4.3.5 (*Interpretation*): Consider the case (4.3.5). Here Proposition 4.3.4 states that the rescaled algorithm (y_n) behaves asymptotically like the process generated by

$$\mathcal{L}f(x) = (1 - p_A - xp_A)f'(x) + p_B s'(1)x[f(x+1) - f(x)].$$

Intuitively, it is more and more likely to play the arm A (the one with the greatest gain probability). Its successes and failures appear within the drift term of the limit infinitesimal generator, whereas playing the arm B with success will provoke a jump. Finally, playing the arm B with failure does not affect the limit dynamics of the process (as $\widetilde{p}_1$ does not appear within the limit generator). To carry out the computations in this section, where we establish the speed of convergence of $(\mathcal{L}_n)$ toward $\mathcal{L}$, the main idea is to condition $\mathbb{E}[y_{n+1}]$ given typical events on the one hand, and rare events on the other hand. Typical events generally construct the drift term of $\mathcal{L}$ and rare events are responsible of the jump term of $\mathcal{L}$ (see also Remark 4.3.3).

Note that one can tune the frequency of jumps with the parameter $s'(1)$. The more concave s is in a neighborhood of 1, the better the convergence is. In particular, if $s'(1) = 0$, the limit process is deterministic. Also, note that choosing a function s non-symmetric with respect to $(1/2, 1/2)$ introduces an a priori bias; see Figure 4.3.4.

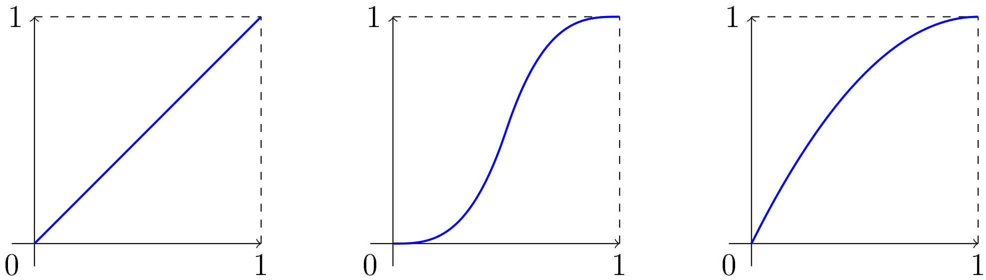


Figure 4.3.4 – Various strategies for $s(x) = x$, s concave, s with a bias

$\diamondsuit$

Let us start the analysis of the rescaled PBA with a global result about a large class of PDMPs, whose proof is postponed to Section 4.5. This lemma provides the necessary arguments to check Assumtion 4.2.3.

Proposition 4.3.6 (*Assumption 4.2.3 for PDMPs*)

Let X be a PDMP with infinitesimal generator

$$\mathcal{L}f(x) = (a - bx)f'(x) + (c + dx)[f(x + 1) - f(x)],$$

such that $a, b, c, d \geq 0$. Assume that either $b > 0$, or $b = 0$ and $a \neq 0$. If $f \in \mathscr{C}_b^N$, then, for all $0 \leq t \leq T$, $P_t f \in \mathscr{C}_b^N$. Moreover, for all $n \leq N$,

$$\|(P_t f)^{(n)}\|_\infty \leq \begin{cases} \sum_{k=0}^n \left(\frac{2|d|}{b}\right)^{n-k} \|f^{(k)}\|_\infty & \text{if } b > 0 \\ \sum_{k=0}^n \frac{n!}{k!}(2|d|T)^{n-k}\|f^{(k)}\|_\infty & \text{if } b = 0 \end{cases}.$$

Note that a very similar result is obtained in [BR15a], but for PDMPs with a diffusive component.

Remark 4.3.7 (*The stationary probability distribution*): Let $(X_t)_{t\geq 0}$ be the PDMP generated by $\mathcal{L}$ defined in Proposition 4.3.6. By using the same tools as in [LP08b, Theorem 6], it is possible to prove existence and uniqueness of a stationary distribution π on $\mathbb{R}_+$. Applying Dynkin's formula with $f(x) = x$, we get

$$\partial_t \mathbb{E}[X_t] = a + c - (b-d)\mathbb{E}[X_t].$$

If one uses the same technique with $f(x) = x^n$, it is possible to deduce the n^{th} moment of the invariant measure π, and Dynkin's formula applied to $f(x) = \exp(\lambda x)$ provides exponential moments of π (see [BMP+15, Remark 2.2] for the detail).

In the setting of (4.3.6), one can use the reasoning above to show that, by denoting by $m_n = \int_0^\infty x^n \pi(dx)$ for $n \geq 0$,

$$m_n = \frac{-p_0'(1)}{n(p_1(1)+p_0'(1))}\sum_{k=1}^{n-2}\binom{n}{k-1}m_k + \frac{2\widetilde{p}_0(1)+(n-1)p_0'(1)}{2(p_1(1)+p_0'(1))}m_{n-1},$$

with the convention $\sum_{k=i+1}^{i} = 0$. $\diamondsuit$

Proof of Proposition 4.3.4: First, let us specify the announced convergence of $\mathcal{L}_n$ toward $\mathcal{L}$; recall that $\gamma_n = n^{-1/2}$ and $\chi_d(y) = \sum_{k=0}^d |y|^k$, so that $I_n(y)$ in (4.3.7) rewrites

$$I_n(y) = \begin{cases} \frac{\sqrt{n+1}-\sqrt{n}-1}{\sqrt{n}}y & \text{with probability } p_1(1-\gamma_n y) \\ 1 + \frac{\sqrt{n+1}-\sqrt{n}-1}{\sqrt{n}}y & \text{with probability } p_0(1-\gamma_n y) \\ \frac{\sqrt{n}-\sqrt{n+1}}{\sqrt{n+1}}y & \text{with probability } \widetilde{p}_1(1-\gamma_n y) \\ \frac{1}{\sqrt{n+1}} + \frac{\sqrt{n}-\sqrt{n+1}}{\sqrt{n+1}}y & \text{with probability } \widetilde{p}_0(1-\gamma_n y) \end{cases},$$

and the infinitesimal generator rewrites

$$\begin{aligned}\mathcal{L}_n f(y) = &\frac{p_1(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+I_n^1(y)\right)-f(y)\right] + \frac{p_0(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+I_n^0(y)\right)-f(y)\right] \\ &+ \frac{\widetilde{p}_1(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+\widetilde{I}_n^1(y)\right)-f(y)\right] + \frac{\widetilde{p}_0(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+\widetilde{I}_n^0(y)\right)-f(y)\right].\end{aligned} \tag{4.3.8}$$

In the sequel, the Landau notation $\mathcal{O}$ will be deterministic and uniform over both y and f.

First, we consider the first term of (4.3.8) and observe that

$$p_1(1-\gamma_n y) = p_1(1) + y\mathcal{O}(\gamma_n),$$

and that

$$I_n^1(y) = \left(\frac{\gamma_n}{\gamma_{n+1}} - 1 - \gamma_n\right)y = \left(\frac{1}{2n} + o(n^{-1}) - \frac{1}{\sqrt{n}}\right)y = -y\gamma_n(1+\mathcal{O}(\gamma_n)),$$

so that $I_n^1(y)^2 = y^2\mathcal{O}(\gamma_n^2)$. Since $\gamma_n \sim \gamma_{n+1}$, and since the Taylor formula gives a random variable ξ_n^y such that

$$f\left(y+I_n^1(y)\right) - f(y) = I_n^1(y)f'(y) + \frac{I_n^1(y)^2}{2}f''(\xi_n^y),$$

we have

$$\gamma_{n+1}^{-1}\left[f\left(y+I_n^1(y)\right)-f(y)\right]=-yf'(y)+\chi_2(y)(\|f'\|_\infty+\|f''\|_\infty)\mathcal{O}(\gamma_n).$$

Then, easy computations show that

$$\frac{p_1(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+I_n^1(y)\right)-f(y)\right]=-p_1(1)yf'(y)+\chi_3(y)(\|f'\|_\infty+\|f''\|_\infty)\mathcal{O}(\gamma_n). \tag{4.3.9}$$

The third term in (4.3.8) is expanded similarly and writes

$$\frac{\widetilde{p}_1(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+\widetilde{I}_n^1(y)\right)-f(y)\right]=\chi_3(y)(\|f'\|_\infty+\|f''\|_\infty)\mathcal{O}(\gamma_n), \tag{4.3.10}$$

while the fourth term becomes

$$\frac{\widetilde{p}_0(1-\gamma_n y)}{\gamma_{n+1}}\left[f\left(y+\widetilde{I}_n^0(y)\right)-f(y)\right]=\widetilde{p}_0(1)f'(y)+\chi_3(y)(\|f'\|_\infty+\|f''\|_\infty)\mathcal{O}(\gamma_n). \tag{4.3.11}$$

Note the slight difference with the expansion of the second term, since we have, on the one hand,

$$\frac{p_0(1-\gamma_n y)}{\gamma_{n+1}}=-\frac{\gamma_n}{\gamma_{n+1}}yp_0'(1)+\frac{\gamma_n^2}{\gamma_{n+1}}y^2p''(\xi_n^y)=-yp_0'(1)+\chi_2(y)\mathcal{O}(\gamma_n),$$

where ξ_n^y is a random variable, while, on the other hand,

$$f(y+I_n^0(y))-f(y)=f(y+1)-f(y)+\chi_1(y)\|f'\|_\infty\mathcal{O}(\gamma_n).$$

Then,

$$\begin{aligned}\frac{p_0(1-\gamma_n y)}{\gamma_{n+1}}&\left[f\left(y+I_n^0(y)\right)-f(y)\right]=\\ &-yp_0'(1)[f(y+1)-f(y)]+\chi_3(y)(\|f\|_\infty+\|f'\|_\infty)\mathcal{O}(\gamma_n).\end{aligned} \tag{4.3.12}$$

Finally, combining (4.3.9), (4.3.10), (4.3.11) and (4.3.12), we obtain the following speed of convergence for the infinitesimal generators:

$$|\mathcal{L}_nf(y)-\mathcal{L}f(y)|=\chi_3(y)(\|f\|_\infty+\|f'\|_\infty+\|f''\|_\infty)\mathcal{O}(\gamma_n), \tag{4.3.13}$$

establishing that the rescaled PBA satisfies Assumption 4.2.2 with $d_1=3$, $N_1=2$ and $\epsilon_n=\gamma_n$. Assumption 4.2.3 follows from Proposition 4.3.6 with $N_2=2$.

In order to apply Theorem 4.2.7, it would remain to check Assumption 4.2.4, that is to prove that the moments of order 3 of (y_n) are uniformly bounded. This happens to be very difficult and we do not even know whether it is true. As an illustration of this difficulty, the reader may refer to [GPS15, Remark 4.4], where uniform bounds for the first moment are provided using rather technical lemmas, and only for an overpenalized version of the algorithm.

In order to overcome this technical difficulty, we introduce a truncated Markov chain coupled with (y_n), which does satisfy a Lyapunov criterion. For $l\in\mathbb{N}^\star$ and $\delta\in(0,1]$, we define $(y_n^{(l,\delta)})_{n\geq 0}$ as follows:

$$y_n^{(l,\delta)}:=\begin{cases}y_n & \text{for } n\leq l\\ \left(y_{n-1}^{(l,\delta)}+I_{n-1}(y_{n-1}^{(l,\delta)})\right)\wedge\delta\gamma_n^{-1} & \text{for } n>l\end{cases}.$$

In the sequel, we denote with an exposant (l,δ) the equivalents of $\mathcal{L}_n, Y_t, \mu_t$ for $(y_n^{(l,\delta)})_{n\geq 0}$. We prove that $(\mathcal{L}_n^{(l,\delta)})_{n\geq 0}$ satisfies our main assumptions, and consequently $(\mu_t^{(l,\delta)})_{t\geq 0}$ is an asymptotic pseudotrajectory of Φ (at least for δ small enough and l large enough), which is the result of the combination of Lemma 4.3.8 and Theorem 4.2.7.

Lemma 4.3.8 (*Behavior of* $(\mu_t^{(l,\delta)})_{t\geq 0}$)

For δ small enough and l large enough, the inhomogeneous Markov chain $(y_n^{(l,\delta)})_{n\geq 0}$ satisfies Assumptions 4.2.2, 4.2.3, 4.2.4 and *4.2.12.*

Now, we shall prove that $(\mu_t)_{t\geq 0}$ is an asymptotic pseudotrajectory of Φ as well. Indeed, let $\varepsilon > 0$ and l be large enough such that $\mathbb{P}(\forall n \geq l, \gamma_n y_n \leq \delta) \geq 1-\varepsilon$ (it is possible since $\gamma_n y_n = 1 - x_n$ converges to 0 in probability). Then, for $T > 0, f \in \mathscr{F}, s \in [0,T]$

$$\begin{aligned}
|\mu_{t+s}(f) - \Phi(\mu_t, s)(f)| &\leq \left|\mu_{t+s}(f) - \mu_{t+s}^{(l,\delta)}(f)\right| + \left|\Phi(\mu_t^{(l,\delta)}, s)(f) - \Phi(\mu_t, s)(f)\right| \\
&\quad + \left|\mu_{t+s}^{(l,\delta)}(f) - \Phi(\mu_t^{(l,\delta)}, s)(f)\right| \\
&\leq (2\|f\|_\infty + 2\|f\|_\infty)(1 - \mathbb{P}(\forall n \geq l, \gamma_n y_n \leq \delta)) \\
&\quad + \left|\mu_{t+s}^{(l,\delta)}(f) - \Phi(\mu_t^{(l,\delta)}, s)(f)\right| \\
&\leq 4\varepsilon + \left|\mu_{t+s}^{(l,\delta)}(f) - \Phi(\mu_t^{(l,\delta)}, s)(f)\right|,
\end{aligned}$$

since $\|f\|_\infty \leq 1$. Taking the suprema over $[0,T]$ and $\mathscr{F}$ yields

$$\limsup_{t\to\infty} \sup_{s\in[0,T]} d_{\mathscr{F}}(\mu_{t+s}, \Phi(\mu_t, s)) \leq 4\varepsilon + \limsup_{t\to\infty} \sup_{s\in[0,T]} d_{\mathscr{F}}(\mu_{t+s}^{(l,\delta)}, \Phi(\mu_t^{(l,\delta)}, s)). \tag{4.3.14}$$

Using Lemma 4.3.8, Theorem 4.2.7 holds for $(\mu_t^{(l,\delta)})_{t\geq 0}$ and (4.3.14) rewrites

$$\limsup_{t\to\infty} \sup_{s\in[0,T]} d_{\mathscr{F}}(\mu_{t+s}, \Phi(\mu_t, s)) \leq 4\varepsilon,$$

so that $(\mu_t)_{t\geq 0}$ is an asymptotic pseudotrajectory of Φ.

Finally, for $t > 0, T > 0, f \in \mathscr{C}_b^0, s \in [0,T]$, set $\nu_t := \mathscr{L}((Y_s^{(t)})_{0\leq T})$ and $\nu := \mathscr{L}((X_s^\pi)_{0\leq T})$. We have

$$\begin{aligned}
|\nu_t(f) - \nu(f)| &\leq \left|\nu_t(f) - \nu_t^{(l,\delta)}(f)\right| + \left|\nu_t^{(l,\delta)}(f) - \nu(f)\right| \\
&\leq 2\|f\|_\infty(1 - \mathbb{P}(\forall n \geq l, \gamma_n y_n \leq \delta)) + \left|\nu_t^{(l,\delta)}(f) - \nu(f)\right| \\
&\leq 2\varepsilon + \left|\nu_t^{(l,\delta)}(f) - \nu(f)\right|.
\end{aligned} \tag{4.3.15}$$

Since $(y_n^{(l,\delta)})_{n\geq 0}$ satisfies Assumption 4.2.12, we can apply Theorem 4.2.13 so that the right-hand side of (4.3.15) converges to 0, which concludes the proof. □

Remark 4.3.9 (*Rate of convergence toward the stationary measure*): For such PDMPs, exponential convergence in Wasserstein distance has already been obtained (see [BMP+15, Proposition 2.1] or [GPS15, Theorem 3.4]). However, we are not in the setting of Theorem 4.2.9, since $\gamma_n = n^{-1/2}$. Thus, $\lambda(\gamma, \epsilon) = 0$, and there is no exponential convergence. This highlights the fact that the rescaled algorithm converges too slowly toward the limit PDMP. ♢

Remark 4.3.10 (*The overpenalized bandit algorithm*): Even though we do not consider the overpenalized bandit algorithm introduced in [GPS15], the tools are the same. The behavior of this algorithm is the same as the PBA's, except from a possible (random) penalization of an arm in case of a success; it writes

$$x_{n+1} = x_n + \gamma_{n+1}\left(X_{n+1} - x_n\right) + \gamma_{n+1}^2\left(\widetilde{X}_{n+1} - x_n\right),$$

where

$$(X_{n+1}, \widetilde{X}_{n+1}) = \begin{cases} (1, x_n) & \text{with probability } p_A x_n \sigma \\ (0, x_n) & \text{with probability } p_B(1-x_n)\sigma \\ (1, 0) & \text{with probability } p_A x_n(1-\sigma) \\ (0, 1) & \text{with probability } p_B(1-x_n)(1-\sigma) \\ (x_n, 1) & \text{with probability } (1-p_B)(1-x_n) \\ (x_n, 0) & \text{with probability } (1-p_A)x_n \end{cases}.$$

Setting $y_n = \gamma_n^{-1}(1 - x_n)$, and following our previous computations, it is easy to show that the rescaled overpenalized algorithm converges, in the sense of Assumption 4.2.2, toward

$$\mathcal{L}f(y) = [1 - \sigma p_A - p_A y]f'(y) + p_B y[f(y+1) - f(y)].$$

♢

4.3.3 Decreasing Step Euler Scheme

In this section, we turn to the study of the so-called *Decreasing Step Euler Scheme* (DSES). This classical stochastic procedure is designed to approximate the stationary measure of a diffusion process of the form

$$X_t^x = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s \tag{4.3.16}$$

with a discrete Markov chain

$$y_{n+1} := y_n + \gamma_{n+1}b(y_n) + \sqrt{\gamma_{n+1}}\sigma(y_n)E_{n+1}, \tag{4.3.17}$$

for any non-increasing sequence $(\gamma_n)_{n\geq 1}$ converging toward 0 such that $\sum_{n=1}^{\infty}\gamma_n = +\infty$ and (E_n) a suitable sequence of random variables. In the sequel, we shall recover the convergence of the DSES toward the diffusion process at equilibrium, as defined by (4.3.16). If $\gamma_n = \gamma$ in (4.3.17), this model would be a constant step Euler scheme as studied by [Tal84, TT90], which approaches the diffusion process at time t when γ tends to 0. By letting $t \to +\infty$ in (4.3.16), it converges to the equilibrium of the diffusion process. We can concatenate those steps by choosing γ_n vanishing but such that $\sum_n \gamma_n$ diverges. The DSES has already been studied in the literature, see for instance [LP02, Lem05].

It is simple, following the computations of Sections 4.3.1 and 4.3.2, to check that $\mathcal{L}_n$ converges (in the sense of Assumption 4.2.2) toward

$$\mathcal{L}f(y) := b(y)f'(y) + \frac{\sigma^2(y)}{2}f''(y).$$

In the sequel, define $\mathscr{F}$ as in Theorem 4.2.7 with $N_2 = 3$.

Proposition 4.3.11 (*Results for the DSES*)

Assume that (E_n) is a sequence of sub-gaussian random variables (i.e. there exists $\kappa > 0$ such that $\forall\theta \in \mathbb{R}, \mathbb{E}[\exp(\theta E_1)] \leq \exp(\kappa\theta^2/2)$), and $\mathbb{E}[E_1] = 0$ and $\mathbb{E}[E_1^2] = 1$. Moreover, assume that $b, \sigma \in \mathscr{C}^\infty$ whose derivatives of any order are bounded, and that σ is bounded. Eventually, assume that there exist constants $0 < b_1 \leq b_2$ and $0 < \sigma_1$ such that, for $|y| > A$,

$$-b_2y^2 \leq b(y)y \leq -b_1y^2, \quad \sigma_1 \leq \sigma(y). \tag{4.3.18}$$

If $\gamma_n = 1/n$, then (μ_t) is a $\frac{1}{2}$-pseudotrajectory of Φ, with respect to $d_{\mathscr{F}}$.

Moreover, there exists a probability distribution π and $C, u > 0$ such that, for all $t > 0$,

$$d_{\mathscr{F}}(\mu_t, \pi) \leq Ce^{-ut}.$$

Furthermore, the sequence of processes $(Y_s^{(t)})_{s\geq 0}$ converges in distribution, as $t \to +\infty$, toward $(X_s^\pi)_{s\geq 0}$ in the Skorokhod space.

Note that one could choose a more general (γ_n), provided that $\lambda(\gamma, \gamma) > 0$. In contrast to classical results, Proposition 4.3.11 provides functional convergence. Moreover, we obtain a rate of convergence in a more general setting than [Lem05, Theorem IV.1], see also [LP02]. Indeed, let us detail the difference between those settings with the example of the Kolmogorov-Langevin equation:

$$dX_t = \nabla V(X_t)dt + \sigma dB_t.$$

A rate of convergence may be obtained in [Lem05] only for V uniformly convex; although, we only need V to be convex outside some compact set. Let us recall that the uniform convexity is a strong assumption ensuring log-Sobolev inequality, Wassertsein contraction... See for instance [Bak94, ABC+00].

Proof of Proposition 4.3.11: Recalling (y_n) in (4.3.17) and $\mathcal{L}_n$ in (4.1.1), we have

$$\mathcal{L}_n(y) = \gamma_{n+1}^{-1}\mathbb{E}\left[f(y + \gamma_{n+1}b(y) + \sqrt{\gamma_{n+1}}\sigma(y)E_{n+1}) - f(y)|y_n = y\right].$$

Easy computations show that Assumption 4.2.2 holds with $\epsilon_n = \sqrt{\gamma_n}, N_1 = 3, d_1 = 3$.

We aim at proving Assumption 4.2.3, i.e. for $f \in \mathscr{F}, j \leq 3$ and $t \leq T$, that $(P_tf)^{(j)}$ exists and

$$\|(P_tf)^{(j)}\|_\infty \leq C_T \sum_{k=0}^{3} \|f^{(k)}\|_\infty.$$

It is straightforward for $j = 0$, but computations are more involved for $j \geq 1$. Let us denote by $(X_t^x)_{t\geq 0}$ the solution of (4.3.16) starting at x. Since b and σ are smooth with bounded derivatives, it is standard that $x \mapsto X_t^x$ is $\mathscr{C}^4$ (see for instance [Kun84, Chapter II, Theorem 3.3]). Moreover, $\partial_x X_t^x$ satisfies the following SDE:

$$\partial_x X_t^x = 1 + \int_0^t b'(X_s^x)\partial_x X_s^x ds + \int_0^t \sigma'(X_s^x)\partial_x X_s^x dW_s.$$

For our purpose, we need the following lemma, which provides a constant for Assumption 4.2.3 of the form $C_T = C_1 \mathrm{e}^{C_2 T}$. Even though we do not explicit the constants for the second and third derivatives in its proof, it is still possible; the main result of the lemma being that we can check Assumption 4.2.8.ii).

Lemma 4.3.12 (*Estimates for the derivatives of the diffusion*)

Under the assumptions of Proposition 4.3.11, for $p \geq 2$ and $t \leq T$,

$$\mathbb{E}[|\partial_x X_t^x|^p] \leq \exp\left(\left(p\|b'\|_\infty + \frac{p(p-1)}{2}\|\sigma'\|_\infty^2\right)T\right),$$

and

$$\mathbb{E}[|\partial_x X_t^x|] \leq \exp\left(\left(\|b'\|_\infty + \frac{1}{2}\|\sigma'\|_\infty^2\right)T\right).$$

For any $p \in \mathbb{N}^\star$, there exist positive constants C_1, C_2 not depending on x, such that

$$\mathbb{E}[|\partial_x^2 X_t^x|^p] \leq C_1 \mathrm{e}^{C_2 T}, \quad \mathbb{E}[|\partial_x^3 X_t^x|^p] \leq C_1 \mathrm{e}^{C_2 T}.$$

The proof of the lemma is postponed to Section 4.5. Using Lemma 4.3.12, and since f and its derivatives are bounded, it is clear that $x \mapsto P_t f(x)$ is three times differentiable, with

$$\begin{aligned}
(P_t f)'(x) &= \mathbb{E}\Big[f'(X_t^x)\partial_x X_t^x\Big], \\
(P_t f)''(x) &= \mathbb{E}\Big[f''(X_t^x)(\partial_x X_t^x)^2 + f'(X_t^x)(\partial_x^2 X_t^x)\Big], \\
(P_t f)^{(3)}(x) &= \mathbb{E}\Big[f^{(3)}(X_t^x)(\partial_x X_t^x)^3 + 3f''(X_t^x)(\partial_x X_t^x)(\partial_x^2 X_t^x) + f'(X_t^x)(\partial_x^3 X_t^x)\Big].
\end{aligned}$$

As a consequence, Assumption 4.2.3 holds, with $C_T = 3C_1^3 \mathrm{e}^{3C_2 T}$ and $N_2 = 3$.

Now, we shall prove that Assumption 4.2.4.ii) holds with $V(y) = \exp(\theta y)$, for some (small) $\theta > 0$. Thanks to (4.3.18), we easily check that, for $\widetilde{V}(y) = 1 + y^2$,

$$\mathcal{L}\widetilde{V}(y) \leq -\widetilde{\alpha}\widetilde{V}(y) + \widetilde{\beta}, \quad \text{with } \widetilde{\alpha} = 2b_1, \widetilde{\beta} = (2b_1 + S) \vee \left(A \sup_{[-A,A]} b + \frac{S^2}{2} + 2b_1(1 + A^2)\right). \tag{4.3.19}$$

Then, [Lem05, Proposition III.1] entails Assumption 4.2.4.ii). Finally, Theorem 4.2.7 applies and we recover [KY03, Theorem 2.1, Chapter 10].

Then, Theorem 4.2.7 provides the asymptotic behavior of the Markov chain $(y_n)_{n\geq 0}$ (in the sense of asymptotic pseudotrajectories). If furtherly we want speeds of convergence, we shall use Theorem 4.2.9 and prove the ergodicity of the limit process; to that end, combine (4.3.19) with [MT93b, Theorem 6.1] (which provides exponential ergodicity for the diffusion toward some stationary measure π), as well as Lemma 4.3.12, to ensure Assumption 4.2.8.ii) with $\mathscr{G} = \{g \in \mathscr{C}^0(\mathbb{R}) : |g(y)| \leq 1+y^2\}$ (v and r are not explicitly given). Note that we used the fact that σ is lower-bounded, which implies that the compact sets are small sets. Moreover, the choice $\gamma_n = n^{-1}$ implies $\lambda(\gamma, \epsilon) = 1/2$. Then, the assumptions of Theorem 4.2.9 are satisfied, with $u_0 = v(1 + 2v + 2r)^{-1}$.

Finally, we can easily check Assumption 4.2.12 for some $d \in \mathbb{N}$, since y_n admits uniformly bounded exponential moments. Then using Theorem 4.2.13 ends the proof. $\square$

4.3.4 Lazier and Lazier Random Walk

We consider the *Lazier and Lazier Random Walk* (LLRW) $(y_n)_{n\geq 0}$ defined as follows:

$$y_{n+1} := \begin{cases} y_n + Z_{n+1} & \text{with probability } \gamma_{n+1} \\ y_n & \text{with probability } 1 - \gamma_{n+1} \end{cases},$$

where (Z_n) is such that $\mathscr{L}(Z_{n+1}|y_0, \ldots, y_n) = \mathscr{L}(Z_{n+1}|y_n)$; we denote the conditional distribution $Q(y_n, \cdot) := \mathscr{L}(Z_{n+1}|y_n)$. In the sequel, define $\mathscr{F} := \{f \in \mathscr{C}_b^0 : 7\|f\|_\infty \leq 1\}$ and $\mathcal{L}f(y) = \int_{\mathbb{R}} f(y+z)Q(y,dz) - f(y)$, which is the generator of a pure-jump Markov process (constant between two jumps).

This example is very simple and could be studied without using our main results; however, we still develop it in order to check the sharpness of our rates of convergence (see Remak 4.3.14).

Proposition 4.3.13 (*Results for the LLRW model*)

The sequence (μ_t) is an asymptotic pseudotrajectory of Φ, with respect to $d_{\mathscr{F}}$.

Moreover, if $\lambda(\gamma,\gamma) > 0$, then (μ_t) is a $\lambda(\gamma,\gamma)$-pseudotrajectory of Φ.

Furthermore, if $\mathcal{L}$ satisfies Assumption 4.2.8.i) for some $v > 0$ then, for any $u < v \wedge \lambda(\gamma,\gamma)$, there exists a constant C such that, for all $t > 0$,

$$d_{\mathscr{F}}(\mu_t, \pi) \leq C e^{-ut}.$$

Remark that the distance $d_{\mathscr{F}}$ in Proposition 4.3.13 is the total variation distance up to a constant.

Proof of Proposition 4.3.13: It is easy to check that (4.1.1) entails

$$\mathcal{L}_n f(y) = \int_{\mathbb{R}} f(y+z)Q(y,dz) - f(y) = \mathcal{L}f(y).$$

It is clear that the LLRW satisfies Assumption 4.2.2 with $d_1 = 0, N_1 = 0, \epsilon_n = 0$, and Assumption 4.2.3 with $C_T = 1, N_2 = 0$. Since $d = d_1 = 0$, Assumption 4.2.4 is also clearly satisfied. Eventually, note that $\lambda(\gamma, \epsilon) = \lambda(\gamma, \gamma)$. Then, Theorem 4.2.7 holds. Finally, if $\mathcal{L}$ satisfies Assumption 4.2.8.i), it is clear that Theorem 4.2.9 applies. $\square$

The assumption on $\mathcal{L}$ satisfying Assumption 4.2.8.i) (which strongly depends on the choice of Q), can be checked with the help of a Foster-Lyapunov criterion, see [MT93b] for instance.

Remark 4.3.14 (*Speed of convergence under Doeblin condition*): Assume there exists a measure ψ and $\varepsilon > 0$ such that for every y and measurable set A, we have

$$\int \mathbb{1}_{y+z\in A} Q(y, dz) \geq \varepsilon \psi(A).$$

It is the classical Doeblin condition, which ensures exponential uniform ergodicity in total variation distance. It is classic to prove that under this condition there exists an invariant distribution π, such that , for every μ and $t \geq 0$

$$d_{\mathscr{F}}(\mu P_t, \pi) \leq e^{-t\varepsilon} d_{\mathscr{F}}(\mu, \pi) \leq e^{-t\varepsilon}$$

Indeed, one can couple two trajectories as follows: choose the same jump times and, using the Doeblin condition, at each jumps, couple them with probability ε. The coupling time then follows an exponential distribution with parameter ε. Then, the conclusion of Proposition 4.3.13 holds with $v = \varepsilon^{-1}$.

However, one can use the Doeblin argument directly with the inhomogeneous chain. Let us denote by (K_n) its sequence of transition kernels. From the Doeblin condition, we have for every μ, ν and $n \geq 0$

$$d_{\mathscr{F}}(\mu K_n, \nu K_n) \leq (1 - \gamma_{n+1}\varepsilon) d_{\mathscr{F}}(\mu, \nu).$$

and as π is invariant for K_n (it is straighforward because π is invariant for Q) then

$$d_{\mathscr{F}}(\mu K_n, \pi) \leq (1 - \gamma_{n+1}\varepsilon) d_{\mathscr{F}}(\mu, \pi).$$

A recursion argument then gives

$$d_{\mathscr{F}}(\mathcal{L}(y_n), \pi) \leq \prod_{k=0}^{n} (1 - \gamma_{k+1}\varepsilon) d_{\mathscr{F}}(\mathcal{L}(y_0), \pi).$$

But,

$$\prod_{k=0}^{n} (1 - \gamma_{k+1}\varepsilon) = \exp\left(\sum_{k=0}^{n} \ln(1 - \gamma_{k+1}\varepsilon)\right) \leq \exp\left(\sum_{k=0}^{n} \ln(1 - \gamma_{k+1}\varepsilon)\right) \leq e^{-\varepsilon \sum_{k=0}^{n} \gamma_{k+1}}.$$

As a conclusion, Proposition 4.3.13 and the direct approach provide the same rate of convergence. This illustrate that our two step method (convergence to a Markov process which converges to equlibrium) does not heavily alter the speed. $\diamondsuit$

Remark 4.3.15 (*Non-convergence in total variation*): Assume that $y_n \in \mathbb{R}_+$ and $Z_n = -y_n/2$. We then have that

$$y_n = \prod_{i=1}^{n} \widetilde{\Theta}_i y_0, \quad \widetilde{\Theta}_i = \left\{ \begin{array}{ll} 1 & \text{with probability } 1-\gamma_i \\ \frac{1}{2} & \text{with probability } \gamma_i \end{array} \right. .$$

where $\widetilde{\Theta}_i$ are independent random variables. Borel-Cantelli's Lemma entails that $(y_n)_{n\geq 0}$ converges to 0 almost surely and, here,

$$\mathcal{L}f(y) = f\left(\frac{y}{2}\right) - f(y).$$

A process with such a generator never hits 0 whenever it starts with a positive value and, then, does not converge in total variation distance. Nevertheless, it is easy to prove that for any y and $t \geq 0$,

$$d_{\mathcal{G}}(\delta_y P_t, \delta_0) \leq \mathbb{E}\left[\frac{1}{2^{N_t}}\right] y \leq e^{-t/2} y,$$

where $\mathcal{G}$ is any class of functions included in $\{f \in \mathscr{C}_b^1 : \|f'\|_\infty \leq 1\}$, and (N_t) a Poisson process. In particular Assumption 4.2.8.ii) holds and there is convergence of our chain to zero in distribution, as well as a rate of convergence in the Fortet-Mourier distance. $\diamondsuit$

4.4 Proofs of theorems

In the sequel, we consider the following classes of functions:

$$\begin{aligned}
\mathscr{F}_1 &:= \{f \in \mathcal{D}(\mathcal{L}) : \mathcal{L}f \in \mathcal{D}(\mathcal{L}), \|f\|_\infty + \|\mathcal{L}f\|_\infty + \|\mathcal{L}\mathcal{L}f\|_\infty \leq 1\}, \\
\mathscr{F}_2 &:= \left\{ f \in \mathcal{D}(\mathcal{L}) \cap \mathscr{C}_b^{N_2} : \sum_{j=0}^{N_2} \|f^{(j)}\|_\infty \leq 1 \right\}, \\
\mathscr{F} &:= \mathscr{F}_1 \cap \mathscr{F}_2.
\end{aligned}$$

The class $\mathscr{F}_1$ is particularly useful to control $P_t f$ (see Lemma 4.4.1), and the class $\mathscr{F}_2$ enables us to deal with smooth and bounded functions (for the second part of the proof of Theorem 4.2.7). Note that an important feature of $\mathscr{F}$ is that Lemma 4.2.1 holds for $\mathscr{F}_1 \cap \mathscr{F}_2$, so that $\mathscr{F}$ contains $\mathscr{C}_c^\infty$ "up to a constant".

Let us begin with preliminary remarks on the properties of the semigroup (P_t).

Lemma 4.4.1 (*Expansion of $P_t f$*)

Let $f \in \mathscr{F}_1$. Then, for all $t > 0$, $P_t f \in \mathscr{F}_1$ and

$$\sup_{f \in \mathscr{F}_1} \|P_t f - f - t\mathcal{L}f\|_\infty \leq \frac{t^2}{2}.$$

Proof of Lemma 4.4.1: It is clear that $P_t f \in \mathscr{F}_1$, since for all $g \in \mathcal{D}(\mathcal{L})$, $P_t\mathcal{L}g = \mathcal{L}P_t g$ and $\|P_t g\|_\infty \leq \|g\|_\infty$. Now, if $f \in \mathscr{F}_1$, then

$$P_t f = f + \int_0^t P_s\mathcal{L}f ds = f + t\mathcal{L}f + K(f,t),$$

where $K(f,t) = P_t f - f - t\mathcal{L}f$. Using the mean value inequality, we have, for $x \in \mathbb{R}^D$,

$$\begin{aligned} |K(f,t)(x)| &= \left|\int_0^t P_s\mathcal{L}f(x)ds - \mathcal{L}f(x)\right| \leq \int_0^t |P_s\mathcal{L}f(x) - \mathcal{L}f(x)|ds \\ &\leq \int_0^t s\|\mathcal{L}\mathcal{L}f\|_\infty ds \leq \frac{t^2}{2}, \end{aligned}$$

which concludes the proof. □

Proof of Theorem 4.2.7: For every $t \geq 0$, set $K(f,t) := P_t f - f - t\mathcal{L}f$ and recall that $m(t) = \sup\{n \geq 0 \ : \ t \geq \tau_n\}$. Then, we have $Y_{\tau_{m(t)}} = Y_t$ and $\tau_{m(t)} \leq t < \tau_{m(t)+1}$. Let $0 < s < T$. Using the following telescoping sum, we have

$$\begin{aligned} d_{\mathscr{F}}(\mu_{t+s}, \Phi(\mu_t, s)) &= d_{\mathscr{F}}(\mu_{\tau_{m(t+s)}}, \Phi(\mu_{\tau_{m(t)}}, s)) \\ &\leq d_{\mathscr{F}}(\Phi(\mu_{\tau_{m(t)}}, \tau_{m(t+s)} - \tau_{m(t)}), \Phi(\mu_{\tau_{m(t)}}, s)) \\ &\quad + d_{\mathscr{F}}(\mu_{\tau_{m(t+s)}}, \Phi(\mu_{\tau_{m(t)}}, \tau_{m(t+s)} - \tau_{m(t)})) \\ &\leq d_{\mathscr{F}}(\Phi(\mu_{\tau_{m(t)}}, \tau_{m(t+s)} - \tau_{m(t)}), \Phi(\mu_{\tau_{m(t)}}, s)) \\ &\quad + \sum_{k=m(t)}^{m(t+s)-1} d_{\mathscr{F}}\left(\Phi\left(\mu_{\tau_{k+1}}, \sum_{j=k+2}^{m(t+s)}\gamma_j\right), \Phi\left(\mu_{\tau_k}, \sum_{j=k+1}^{m(t+s)}\gamma_j\right)\right), \end{aligned} \tag{4.4.1}$$

with the convention $\sum_{k=i+1}^{i} = 0$. Our aim is now to bound each term of this sum. The first one is the simplest: indeed, we have $s \leq \tau_{m(t+s)+1} - \tau_{m(t)}$, so $s - \gamma_{m(t+s)+1} \leq \tau_{m(t+s)} - \tau_{m(t)}$ and $\tau_{m(t+s)} - \tau_{m(t)} \leq s + \gamma_{m(t)+1}$. Denoting by $u = s \wedge (\tau_{m(t+s)} - \tau_{m(t)})$ and $h = |\tau_{m(t+s)} - \tau_{m(t)} - s|$ we have, by the semigroup property,

$$d_{\mathscr{F}}\left(\Phi(\mu_t, \tau_{m(t+s)} - \tau_{m(t)}), \Phi(\mu_t, s)\right) = d_{\mathscr{F}}\left(\Phi(\Phi(\mu_t, u), h), \Phi(\mu_t, u)\right).$$

From Lemma 4.4.1, we know that for every $f \in \mathscr{F}_1$ and every probability measure ν,

$$|\Phi(\nu, h)(f) - \nu(f)| = |\nu(P_h f - f)| \leq h + \frac{h^2}{2} \leq \frac{3}{2}h,$$

for $h \leq 1$. It is then straightforward that

$$d_{\mathscr{F}}\left(\Phi(\mu_t, \tau_{m(t+s)} - \tau_{m(t)}), \Phi(\mu_t, s)\right) \leq \frac{3}{2}h \leq \frac{3}{2}\gamma_{m(t)+1}. \tag{4.4.2}$$

Now, we provide bounds for the generic term of the telescoping sum in (4.4.1). Let

$f \in \mathscr{F}_1$ and $m(t) \leq k \leq m(t+s)-1$. On the one hand, using Lemma 4.4.1,

$$\begin{aligned}\Phi\left(\mu_{\tau_k}, \sum_{j=k+1}^{m(t+s)} \gamma_j\right)(f) &= \mu_{\tau_k} P_{\sum_{j=k+1}^{m(t+s)} \gamma_j}(f) \\ &= \mu_{\tau_k}(P_{\tau_{m(t+s)}-\tau_{k+1}} f) + \int_0^{\gamma_{k+1}} \mu_{\tau_k}(\mathcal{L} P_{\tau_{m(t+s)}-\tau_{k+1}+u} f) du \\ &= \mu_{\tau_k}(P_{\tau_{m(t+s)}-\tau_{k+1}} f) + \gamma_{k+1} \mu_{\tau_k}(\mathcal{L} P_{\tau_{m(t+s)}-\tau_{k+1}} f) \\ &\quad + K\left(P_{\tau_{m(t+s)}-\tau_{k+1}} f, \gamma_{k+1}\right).\end{aligned}$$

On the other hand,

$$\mu_{\tau_{k+1}}(f) = \mu_{\tau_k}(f) + \gamma_{k+1} \mu_{\tau_k}(\mathcal{L}_k f)$$

so that

$$\begin{aligned}\Phi\left(\mu_{\tau_{k+1}}, \sum_{j=k+2}^{m(t+s)} \gamma_j\right)(f) &= \mu_{\tau_{k+1}}(P_{\tau_{m(t+s)}-\tau_{k+1}} f) \\ &= \mu_{\tau_k}(P_{\tau_{m(t+s)}-\tau_{k+1}} f) + \gamma_{k+1} \mu_{\tau_k}(\mathcal{L}_k P_{\tau_{m(t+s)}-\tau_{k+1}} f).\end{aligned}$$

Henceforth,

$$\begin{aligned}\Phi\left(\mu_{\tau_{k+1}}, \sum_{j=k+2}^{m(t+s)} \gamma_j\right)(f) - \Phi\left(\mu_{\tau_k}, \sum_{j=k+1}^{m(t+s)} \gamma_j\right)(f) &\leq \gamma_{k+1} \mu_{\tau_k}((\mathcal{L}_k - \mathcal{L}) P_{\tau_{m(t+s)}-\tau_{k+1}} f) \\ &\quad + K\left(P_{\tau_{m(t+s)}-\tau_{k+1}} f, \gamma_{k+1}\right).\end{aligned}$$

Now, we bound the previous term using Assumption 4.2.2, Assumption 4.2.3, and Assumption 4.2.4. Let $m(t) \leq k \leq m(t+s)-1$. Recall that, since $s < T$, $\tau_{m(t+s)} - \tau_{k+1} \leq \tau_{m(t+s)} - \tau_{m(t)+1} \leq (t+s) - t \leq T$. Then, for all $f \in \mathscr{F}_2$,

$$\begin{aligned}&|\mu_{\tau_k}((\mathcal{L}_k - \mathcal{L}) P_{\tau_{m(t+s)}-\tau_{k+1}} f)| \leq \mu_{\tau_k}(|(\mathcal{L}_k - \mathcal{L}) P_{\tau_{m(t+s)}-\tau_{k+1}} f|) \\ &\quad \leq \mu_{\tau_k}\left(M_1 \chi_{d_1} \sum_{j=0}^{N_1} \|(P_{\tau_{m(t+s)}-\tau_{k+1}} f)^{(j)}\|_\infty \epsilon_k\right) \leq \mu_{\tau_k}\left(M_1(N_1+1) C_T \chi_d \sum_{j=0}^{N_2} \|f^{(j)}\|_\infty \epsilon_k\right) \\ &\quad \leq M_1(N_1+1) C_T \mathbb{E}[\chi_d(y_k)] \sum_{j=0}^{N_2} \|f^{(j)}\|_\infty \epsilon_k \leq M_1 M_2 (N_1+1) C_T \sum_{j=0}^{N_2} \|f^{(j)}\|_\infty \epsilon_k \\ &\quad \leq M_1 M_2 (N_1+1) C_T \epsilon_k.\end{aligned}$$

Gathering the previous bounds entails

$$\begin{aligned}&\sum_{k=m(t)}^{m(t+s)-1} d_{\mathscr{F}}\left(\Phi\left(\mu_{\tau_{k+1}}, \sum_{j=k+2}^{m(t+s)} \gamma_j\right), \Phi\left(\mu_{\tau_k}, \sum_{j=k+1}^{m(t+s)} \gamma_j\right)\right) \\ &\quad \leq \sum_{k=m(t)}^{m(t+s)-1} \left(M_1 M_2 (N_1+1) C_T \gamma_{k+1} \epsilon_k + \frac{\gamma_{k+1}^2}{2}\right) \\ &\quad \leq (T+1)\left(M_1 M_2 (N_1+1) C_T + \frac{1}{2}\right)(\gamma_{m(t)} \vee \epsilon_{m(t)}). \qquad (4.4.3)\end{aligned}$$

Thus, combining (4.4.1), (4.4.2) and (4.4.3) yields

$$\sup_{s\le T} d_{\mathscr{F}}(\mu_{t+s},\Phi(\mu_t,s)) \le C'_T(\gamma_{m(t)}\vee\epsilon_{m(t)}), \tag{4.4.4}$$

with $C'_T=\frac{3}{2}+(T+1)\left(M_1M_2(N_1+1)C_T+\frac{1}{2}\right)$. Then, $(\mu_t)_{t\ge 0}$ is an asymptotic pseudotrajectory of Φ (with respect to $d_{\mathscr{F}}$).

Now, we turn to the study of the case $\lambda(\gamma,\epsilon)>0$. For any $\lambda<\lambda(\gamma,\epsilon)$, we have (for n large enough) $\gamma_n\vee\epsilon_n\le\exp(-\lambda\tau_n)$. Then, for any t large enough,

$$\gamma_{m(t)}\vee\epsilon_{m(t)}\le \mathrm{e}^{-\lambda\tau_{m(t)}}\le \mathrm{e}^{\lambda(t-\tau_{m(t)})}\mathrm{e}^{-\lambda t}\le \mathrm{e}^{\lambda(\gamma,\epsilon)}\mathrm{e}^{-\lambda t}.$$

Now, plugging this upper bound in (4.4.4), we get, for $\lambda<\lambda(\gamma,\epsilon)$,

$$\sup_{s\le T} d_{\mathscr{F}}(\mu_{t+s},\Phi(\mu_t,s))\le \mathrm{e}^{\lambda(\gamma,\epsilon)}C'_T\mathrm{e}^{-\lambda t}. \tag{4.4.5}$$

Finally, we can deduce that

$$\limsup_{t\to+\infty}\frac{1}{t}\log\left(\sup_{0\le s\le T} d(\mu_{t+s},\Phi(\mu_t,s))\right)\le-\lambda$$

for any $\lambda<\lambda(\gamma,\epsilon)$, which concludes the proof of Theorem 4.2.7. □

Proof of Theorem 4.2.9: The first part of the proof is an adaptation of [Ben99]. Assume Assumption 4.2.8.i) and, without loss of generality, assume $M_3>1$. If $v>\lambda(\gamma,\epsilon)$, fix $\varepsilon>v-\lambda(\gamma,\epsilon)$, otherwise let $\varepsilon>0$, and set $u:=v-\varepsilon$, $T_\varepsilon:=\varepsilon^{-1}\log M_3$. Since $u<\lambda(\gamma,\epsilon)$, and using (4.4.5), the following sequence of inequalities holds, for any $T\in[T_\varepsilon,2T_\varepsilon]$ and $n\in\mathbb{N}$:

$$\begin{aligned} d_{\mathscr{G}}\left(\mu_{(n+1)T},\pi\right) &\le d_{\mathscr{G}}\left(\mu_{(n+1)T},\Phi(\mu_{nT},T)\right)+d_{\mathscr{G}}\left(\Phi(\mu_{nT},T),\pi\right)\\ &\le \mathrm{e}^{\lambda(\gamma,\epsilon)}C'_T\mathrm{e}^{-unT}+M_3 d_{\mathscr{G}}\left(\mu_{nT},\pi\right)\mathrm{e}^{-vT}\\ &\le \mathrm{e}^{\lambda(\gamma,\epsilon)}C'_T\mathrm{e}^{-unT}+d_{\mathscr{G}}\left(\mu_{nT},\pi\right)\mathrm{e}^{-uT}, \end{aligned}$$

with $C'_T=\frac{3}{2}+(T+1)\left(M_1M_2(N_1+1)C_T+\frac{1}{2}\right)$. Denoting by $\delta_n:=d_{\mathscr{G}}\left(\mu_{nT},\pi\right)$ and $\rho:=\mathrm{e}^{-uT}$, the previous inequality turns into $\delta_{n+1}\le\mathrm{e}^{\lambda(\gamma,\epsilon)}C'_T\rho^n+\rho\delta_n$, from which we derive

$$\delta_n\le n\rho^{n-1}C'_T\mathrm{e}^{\lambda(\gamma,\epsilon)}+\rho^n\delta_0.$$

Hence, for every $n\ge 0$ and $T\in[T_\varepsilon,2T_\varepsilon]$, we have

$$d_{\mathscr{G}}\left(\mu_{nT},\pi\right)\le\mathrm{e}^{-(u-\varepsilon)nT}\left(M_5+d_{\mathscr{G}}\left(\mu_0,\pi\right)\right),\quad M_5=\mathrm{e}^{\lambda(\gamma,\epsilon)}\left(\sup_{n\ge 0} n\mathrm{e}^{-\varepsilon nT}\right)\left(\sup_{T\in[T_\varepsilon,2T_\varepsilon]} C'_T\right).$$

Then, for any $t>T_\varepsilon$, let $n=\lfloor tT_\varepsilon^{-1}\rfloor$ and $T=tn^{-1}$. Then, $T\in[T_\varepsilon,2T_\varepsilon]$ and the following upper bound holds:

$$d_{\mathscr{G}}\left(\mu_t,\pi\right)\le\left(M_5+d_{\mathscr{G}}\left(\mu_0,\pi\right)\right)\mathrm{e}^{-(u-\varepsilon)t}.$$

Now, assume Assumption 4.2.8.ii). For any (small) $\varepsilon > 0$, there exists $e^{\lambda(\gamma,\epsilon)}$ such that $\gamma_{m(t)} \vee \epsilon_{m(t)} \leq e^{\lambda(\gamma,\epsilon)} \exp(-(\lambda(\gamma,\epsilon)-\varepsilon)t)$. For any $\alpha \in (0,1)$, we have

$$\begin{aligned} d_{\mathscr{F}\cap\mathscr{G}}(\mu_t,\pi) &\leq d_{\mathscr{F}\cap\mathscr{G}}(\mu_t,\Phi(\mu_{\alpha t},(1-\alpha)t)) + d_{\mathscr{F}\cap\mathscr{G}}(\Phi(\mu_{\alpha t},(1-\alpha)t),\pi) \\ &\leq C'_{(1-\alpha)t}(\gamma_{m(\alpha t)} \vee \epsilon_{m(\alpha t)}) + M_3 e^{-v(1-\alpha)t} \\ &\leq M_4 e^{r(1-\alpha)t} e^{\lambda(\gamma,\epsilon)} e^{-(\lambda(\gamma,\epsilon)-\varepsilon)\alpha t} + M_3 e^{-v(1-\alpha)t}. \end{aligned} \tag{4.4.6}$$

Optimizing (4.4.6) by taking $\alpha = (r+v)(r+v+\lambda(\gamma,\epsilon)-\varepsilon)^{-1}$, we get

$$d_{\mathscr{F}\cap\mathscr{G}}(\mu_t,\pi) \leq M_5 \exp\left(-\frac{v(\lambda(\gamma,\epsilon)-\varepsilon)}{r+v+\lambda(\gamma,\epsilon)-\varepsilon}t\right),$$

with $M_5 = M_4 e^{\lambda(\gamma,\epsilon)} + M_3$, which depends on ε only through M_3.

Lastly, assume Assumption 4.2.2.iii). Denote by $\mathcal{K}$ the set of probability measures ν such that

$$\nu(W) < M = \sup_{n\geq 0} \mathbb{E}[W(y_n)].$$

Let $\varepsilon > 0$ and $K = \{x \in \mathbb{R}^D : W(x) \leq M/\varepsilon\}$. For every $\nu \in \mathcal{K}$, using Markov's inequality, it is clear that

$$\nu(K^C) \leq \frac{\varepsilon}{M}\nu(W) \leq \varepsilon.$$

Then $\mathcal{K}$ is a relatively compact set (by Prokhorov's Theorem). In the sense of [Ben99], the measure π is an attractor and, since for any $t > 0, \mu_t \in \mathcal{K}$, we can apply [Ben99, Theorem 6.10] to achieve the proof. □

Proof of Theorem 4.2.13: We shall prove the convergence of the sequence of processes $(Y_s^{(t)})_{0\leq s\leq T}$, as $t \to +\infty$, toward $(X_s^{\pi})_{0\leq s\leq T}$ in the Skorokhod space $D([0,T])$, for any $T > 0$. Then, using [Bil99, Theorem 16.7], this convergence entails Theorem 4.2.13, i.e. convergence of the sequence $(Y^{(t)})$ in $D([0,\infty))$.

Let $T > 0$. The proof of functional convergence classically relies on proving the convergence of finite-dimensional distributions, on the one hand, and tightness, on the other hand. First, we prove the former, which is the first part of Theorem 4.2.13. We choose to prove the convergence of the finite-dimensional distributions in the case $m = 2$. The proof for the general case is similar but with a laborious notation. Denote by $T_{u,v}g(y) := \mathbb{E}[g(Y_v)|Y_u = y]$. With this notation, (4.4.4) becomes

$$\sup_{s\leq T}\sup_{g\in\mathscr{F}} (\mu_t T_{t,t+s} g - \mu_t P_s g) \leq C'_T(\gamma_{m(t)} \vee \epsilon_{m(t)}).$$

This upper bound does not depend on μ_t, so, for any probability distribution ν, we have

$$\sup_{s\leq T}\sup_{g\in\mathscr{F}} (\nu T_{t,t+s} g - \nu P_s g) \leq C'_T(\gamma_{m(t)} \vee \epsilon_{m(t)}).$$

This inequality implies that, for any ν,

$$\sup_{s_1\leq s_2\leq T}\sup_{g\in\mathscr{F}} (\nu T_{t+s_1,t+s_2} g - \nu P_{s_2-s_1} g) \leq C'_T(\gamma_{m(t)} \vee \epsilon_{m(t)}), \tag{4.4.7}$$

which converges toward 0 as $t \to +\infty$. From now on, we denote, for any function f, $\widehat{f}_x(y) := f(x,y)$. If f is a smooth function (say in $\mathscr{C}_c^\infty$ with enough derivatives bounded), $\hat{f}.(\cdot) \in \mathscr{F}$. On the one hand, for $0, s_1 < s_2 < T$,

$$\mathbb{E}[f(X_{s_1}^\pi, X_{s_2}^\pi)] = \int P_{s_2-s_1}\widehat{f}_y(y)\pi(dy) = \pi P_{s_2-s_1}\widehat{f}.(\cdot).$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}[f(Y_{s_1}^{(t)}, Y_{s_2}^{(t)}] &= \mathbb{E}\left[\mathbb{E}[f(Y_{s_1}^{(t)}, Y_{s_2}^{(t)}|Y_{s_1}^{(t)}]\right] = \mathbb{E}\left[T_{t+s_1,t+s_2}\widehat{f}_{Y_{t+s_1}}(Y_{t+s_1})\right]\\ &= T_{0,t+s_1}\left(T_{t+s_1,t+s_2}\widehat{f}.(\cdot)\right).\end{aligned}$$

We have the following triangle inequality:

$$\begin{aligned}\left|\mathbb{E}[f(Y_{s_1}^{(t)}, Y_{s_2}^{(t)}] - \mathbb{E}[f(X_{s_1}^\pi, X_{s_2}^\pi)]\right| &= \left|T_{0,t+s_1}\left(T_{t+s_1,t+s_2}\widehat{f}.(\cdot)\right) - \pi P_{s_2-s_1}\widehat{f}.(\cdot)\right|\\ &\leq \left|T_{0,t+s_1}\left(T_{t+s_1,t+s_2}\widehat{f}.(\cdot) - P_{s_2-s_1}\widehat{f}.(\cdot)\right)\right|\\ &\quad + \left|T_{0,t+s_1}\left(P_{s_2-s_1}\widehat{f}.(\cdot)\right) - \pi P_{s_2-s_1}\widehat{f}.(\cdot)\right| \qquad (4.4.8)\end{aligned}$$

Firstly, using (4.4.7), and if $\widehat{f}.(\cdot) \in \mathscr{F}$,

$$\lim_{t\to\infty} T_{0,t+s_1}\left(T_{t+s_1,t+s_2}\widehat{f}.(\cdot) - P_{s_2-s_1}\widehat{f}.(\cdot)\right) = \lim_{t\to\infty} \mu_{t+s_1}\left(T_{t+s_1,t+s_2}\widehat{f}.(\cdot) - P_{s_2-s_1}\widehat{f}.(\cdot)\right) = 0.$$

Secondly, $P_{s_2-s_1}f.(\cdot) \in \mathscr{C}_b^0$ and, using Theorem 4.2.9,

$$\lim_{t\to\infty} T_{0,t+s_1}\left(P_{s_2-s_1}\widehat{f}.(\cdot)\right) - \pi P_{s_2-s_1}\widehat{f}.(\cdot) = 0.$$

From (4.4.8), it is straightforward that, for a smooth f,

$$\lim_{t\to\infty}\left|\mathbb{E}[f(Y_{s_1}^{(t)}, Y_{s_2}^{(t)}] - \mathbb{E}[f(X_{s_1}^\pi, X_{s_2}^\pi)]\right| = 0,$$

and applying Lemma 4.2.1 achieves the proof of finite dimensional convergence for $m = 2$.

To prove tightness, which is the second part of Theorem 4.2.13, we need the following lemma, whose proof is postponed to Section 4.5.

Lemma 4.4.2 (*Martingale properties*)

Let f be a continuous and bounded function. The process $(\widehat{M}_n^f)_{n\geq 0}$, defined for every $n \geq 0$ by

$$\widehat{M}_n^f = f(y_n) - f(y_0) - \sum_{k=0}^{n-1}\gamma_{k+1}\mathcal{L}_k f(y_k),$$

is a martingale, with

$$\langle\widehat{M}^f\rangle_n = \sum_{k=0}^{n-1}\gamma_{k+1}\Gamma_k f(y_k).$$

Moreover, under Assumption 4.2.12, if $d \geq d_2$ then for every $N \geq 0$, there exist a constant $M_7 > 0$ (depending on N and y_0) such that

$$\mathbb{E}\left[\sup_{n \leq N} \chi_{d_1}(y_n)\right] \leq M_7.$$

Now, define

$$M_s^{(t,i)} = \widehat{M}^{\varphi_i}_{m(t+s)} - \widehat{M}^{\varphi_i}_{m(t)},$$

$$A_s^{(t,i)} = \varphi_i(Y_t) + \int_{\tau_{m(t)}}^{\tau_{m(t+s)}} \mathcal{L}_{m(u)}\varphi_i(Y_u)du = \varphi_i(y_{m(t)}) + \sum_{k=m(t)}^{m(t+s)-1} \gamma_{k+1}\mathcal{L}_k\varphi_i(y_k)$$

and

$$Y_s^{(t,i)} = \varphi_i(Y_s^{(t)}).$$

With this notation and Lemma 4.4.2, we have

$$Y_s^{(t,i)} = A_s^{(t,i)} + M_s^{(t,i)}$$

and $(M_s^{(t,i)})_{s\geq 0}$ is a martingale with quadratic variation

$$\langle M^{(t,i)}\rangle_s = \int_{\tau_{m(t)}}^{\tau_{m(t+s)}} \Gamma_{m(u)}\varphi_i(Y_u)du,$$

where Γ_n is as in Assumption 4.2.12. From the convergence of finite-dimensional distributions, for every $s \in [0,T]$, the sequence $(Y_s^{(t)})_{t\geq 0}$ is tight. It is then enough, from the Aldous-Rebolledo criterion (see Theorems 2.2.2 and 2.3.2 in [JM86]) and Lemma 4.4.2 to show that: for every $S \geq 0$, $\varepsilon, \eta > 0$, there exists a $\delta > 0$ and $t_0 > 0$ with the property that whatever the family of stopping times $(\sigma^{(t)})_{t\geq 0}$, with $\sigma^{(t)} \leq S$, for every $i \in \{1, \dots D\}$,

$$\sup_{t\geq t_0}\sup_{\theta\leq\delta} \mathbb{P}\left(\left|\langle M^{(t,i)}\rangle_{\sigma^{(t)}} - \langle M^{(t,i)}\rangle_{\sigma^{(t)}+\theta}\right| \geq \eta\right) \leq \varepsilon \tag{4.4.9}$$

and

$$\sup_{t\geq t_0}\sup_{\theta\leq\delta} \mathbb{P}\left(\left|A^{(t,i)}_{\sigma^{(t)}} - A^{(t,i)}_{\sigma^{(t)}+\theta}\right| \geq \eta\right) \leq \varepsilon. \tag{4.4.10}$$

We have, using Assumption 4.2.12,

$$\begin{aligned} A^{(t,i)}_{\sigma^{(t)}+\theta} - A^{(t,i)}_{\sigma^{(t)}} &= \int_{\tau_{m(t+\sigma^{(t)})}}^{\tau_{m(t+\sigma^{(t)}+\theta)}} \mathcal{L}_{m(u)}\varphi_i(Y_u)du \leq \int_{\tau_{m(t+\sigma^{(t)})}}^{\tau_{m(t+\sigma^{(t)}+\theta)}} M_6\chi_{d_2}(Y_u)du \\ &\leq M_6|\tau_{m(t+\sigma^{(t)}+\theta)} - \tau_{m(t+\sigma^{(t)})}| \sup_{r\leq T}\chi_{d_2}(Y_r). \end{aligned}$$

From the definition of τ_n,

$$|\tau_{m(t+\sigma^{(t)}+\theta)} - \tau_{m(t+\sigma^{(t)})}| \leq \theta + \gamma_{m(t)+1},$$

and then, using Lemma 4.4.2 and Markov's inequality

$$\mathbb{P}\left(\left|A^{(t,i)}_{\sigma^{(t)}} - A^{(t,i)}_{\sigma^{(t)}+\theta}\right| \geq \eta\right) \leq \frac{M_6(\theta+\gamma_{m(t_0)+1})}{\eta}\mathbb{E}[\sup_{s\leq T}\chi_{d_2}(Y_r)] \leq M_6M_7\frac{(\delta+\gamma_{m(t_0)+1})}{\eta}.$$

Proving the inequality (4.4.9) is done in a similar way, and achieves the proof. $\square$

4.5 Appendix

4.5.1 General appendix

Proof of Lemma 4.2.1: Let $f \in \mathscr{C}_b^0, g \in \mathscr{C}_c^\infty$. Note that $fg \in \mathscr{C}_c^0$ and, using Weierstrass' Theorem, it is well known that, for all $\varepsilon > 0$, there exists $\varphi \in \mathscr{C}_c^\infty$ such that $\|fg - \varphi\|_\infty \leq \varepsilon$. By hypothesis, and since $\mathscr{F}$ is a star domain, there exists $\lambda > 0$ such that $\lambda g, \lambda\varphi \in \mathscr{F}$. Then,

$$|\mu_n(fg) - \mu(fg)| \leq |\mu_n(fg) - \mu_n(\varphi)| + \frac{1}{\lambda}|\mu_n(\lambda\varphi) - \mu(\lambda\varphi)| + |\mu(fg) - \mu(\varphi)|,$$

thus $\limsup_{n\to\infty} |\mu_n(fg) - \mu(fg)| \leq 2\varepsilon$. Now,

$$\begin{aligned}
|\mu_n(f) - \mu(f)| &\leq |\mu_n(f - fg) - \mu(f - fg)| + |\mu_n(fg) - \mu(fg)| \\
&\leq \|f\|_\infty |\mu_n(1 - g) - \mu(1 - g)| + |\mu_n(fg) - \mu(fg)| \\
&\leq \frac{\|f\|_\infty}{\lambda} |\mu_n(\lambda g) - \mu(\lambda g)| + |\mu_n(fg) - \mu(fg)|
\end{aligned}$$

so that $\limsup_{n\to\infty} |\mu_n(f) - \mu(f)| \leq 2\varepsilon$, for any $\varepsilon > 0$, which concludes the proof.

Now, assuming $\mathscr{F} \subseteq \mathscr{C}_b^1$, use [Che04, Theorem 5.6]. Then, convergence with respect to $d_\mathscr{F}$ is equivalent to weak convergence. Indeed, $d_{\mathscr{C}_b^1}$ is the well-known Fortet-Mourier distance, which metrizes the weak topology. It is also the Wasserstein distance W_δ, with respect to the distance δ such that

$$\forall x, y \in \mathbb{R}^D, \quad \delta(x, y) = \sup_{f \in \mathscr{C}_b^1} |f(x) - f(y)| = |x - y| \wedge 2.$$

See also [RKSF13, Theorem 4.4.2.]. □

Proof of Lemma 4.4.2: Let $\mathscr{F}_n = \sigma(y_0, \dots, y_n)$ be the natural filtration. Classically, we have

$$\begin{aligned}
\mathbb{E}[\widehat{M}_{n+1}^f \mid \mathscr{F}_n] &= \mathbb{E}[f(y_{n+1}) - f(y_0) - \sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k) \mid \mathscr{F}_n] \\
&= f(y_n) + \gamma_{n+1}\mathcal{L}_n f(y_n) - f(y_0) - \sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k) \\
&= \widehat{M}_n^f.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{E}[(\widehat{M}^f_{n+1})^2 \mid \mathscr{F}_n] &= \mathbb{E}\left[f(y_{n+1})^2 + f(y_0)^2 + \left(\sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right)^2 \,\middle|\, \mathscr{F}_n \right] \\
&\quad - \mathbb{E}\left[2f(y_{n+1}) \left(f(y_0) + \sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right) \,\middle|\, \mathscr{F}_n \right] \\
&\quad + \mathbb{E}\left[2f(y_0) \left(\sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right) \,\middle|\, \mathscr{F}_n \right] \\
&= f(y_n)^2 + \gamma_{n+1}\mathcal{L}_n f^2(y_n) + f(y_0)^2 + \left(\sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right)^2 \\
&\quad - 2(f(y_n) + \gamma_{n+1}\mathcal{L}_n f(y_n)) \left(f(y_0) + \sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right) \\
&\quad + 2f(y_0) \left(\sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right).
\end{aligned}$$

Henceforth,

$$\begin{aligned}
\mathbb{E}[(\widehat{M}^f_{n+1})^2 \mid \mathscr{F}_n] &= \gamma_{n+1}\mathcal{L}_n f^2(y_n) + 2\gamma_{n+1}\mathcal{L}_n f(y_n) \left(\sum_{k=0}^{n-1} \gamma_{k+1}\mathcal{L}_k f(y_k)\right) + (\gamma_{n+1}\mathcal{L}_n f(y_n))^2 \\
&\quad - 2f(y_n)\gamma_{n+1}\mathcal{L}_n f(y_n) - 2\gamma_{n+1}\mathcal{L}_n f(y_n) \left(f(y_0) + \sum_{k=0}^{n} \gamma_{k+1}\mathcal{L}_k f(y_k)\right) \\
&\quad + 2f(y_0)\gamma_{n+1}\mathcal{L}_n f(y_n) + (m^f_n)^2 \\
&= (\widehat{M}^f_n)^2 + \gamma_{n+1}\mathcal{L}_n f^2(y_n) - (\gamma_{n+1}\mathcal{L}_n f(y_n))^2 - 2f(y_n)\gamma_{n+1}\mathcal{L}_n f(y_n) \\
&= (\widehat{M}^f_n)^2 + \gamma_{n+1}\Gamma_n f.
\end{aligned}$$

Now, on the first hand, using Assumption 4.2.12,

$$\mathbb{E}\left[\langle \widehat{M}^{\chi_{d_2}}\rangle_N\right] = \mathbb{E}\left[\sum_{k=0}^{N-1} \gamma_{k+1}\Gamma_{k+1}\chi_{d_2}(y_k)\right] \le M_6 \sum_{k=0}^{N-1} \gamma_{k+1}\mathbb{E}\left[\chi_d(y_k)\right] \le M_2 M_6 \sum_{k=0}^{N-1} \gamma_{k+1},$$

and then Doob's inequality gives

$$\mathbb{E}\left[\left(\sup_{n\le N} \widehat{M}^{\chi_{d_2}}_n\right)^2\right]^{1/2} \le 2\mathbb{E}\left[\langle \widehat{M}^{\chi_{d_2}}\rangle_N\right]^{1/2} \le C,$$

for some constant C, only depending on N. On the other hand, from Lemma 4.4.2 and Assumption 4.2.12,

$$\sup_{n\le N} \chi_{d_2}(y_n) \le \chi_{d_2}(y_0) + M_6 \sum_{k=0}^{N-1} \gamma_{k+1} \sup_{n\le k} \chi_{d_2}(y_n) + \sup_{n\le N} \widehat{M}^{\chi_{d_2}}_n.$$

Using the triangle inequality, we then have

$$\mathbb{E}\left[\left(\sup_{n\le N}\chi_{d_2}(y_n)\right)^2\right]^{1/2} \le \mathbb{E}\left[(\chi_{d_2}(y_0))^2\right]^{1/2} + M_6\sum_{k=0}^{N-1}\gamma_{k+1}\mathbb{E}\left[\left(\sup_{n\le k}\chi_{d_2}(y_n)\right)^2\right]^{1/2}$$
$$+\mathbb{E}\left[\left(\sup_{n\le N}\widehat{M}_n^{\chi_{d_2}}\right)^2\right]^{1/2}.$$

Then, using (discrete) Grönwall's Lemma as well as Cauchy-Schwarz's inequality ends the proof. □

4.5.2 Appendix for the penalized bandit algorithm

Proof of Proposition 4.3.6: The unique solution of the ODE $y'(t) = a - by(t)$ with initial condition x is given by

$$\Psi(x,t) = \begin{cases} \left(x-\frac{a}{b}\right)\mathrm{e}^{-bt}+\frac{a}{b} & \text{if } b>0 \\ x+at & \text{if } b=0 \end{cases}.$$

Firstly, assume that $b>0$ and let $t\in[0,T]$. We have, for $x>0$

$$\begin{aligned}P_tf(x) &= \mathbb{E}_x\left[f(X_t)\right] = f\left(\Psi(x,t)\right)\mathbb{P}_x\left(T>t\right)+\mathbb{E}_x\left[f(X_t)|T\le t\right]\mathbb{P}_x\left(T\le t\right)\\ &= f\left(\Psi(x,t)\right)\exp\left(-\int_0^t(c+d\Psi(x,s))ds\right)\\ &+\int_0^t P_{t-u}f(\Psi(x,u)+1)(c+d\Psi(x,u))\exp\left(-\int_0^u(c+d\Psi(x,s))ds\right)du. \end{aligned}\tag{4.5.1}$$

At this stage, the smoothness of the right-hand side of (4.5.1) with respect to x is not clear. Let $0<\varepsilon<\min(a/b,1/2)$. If $0\le x\le a/b-\varepsilon$, use the substitution

$$v=\Psi(x,u),\quad u=\varphi(x,v)=\frac{1}{b}\log\left(\frac{x-\frac{a}{b}}{v-\frac{a}{b}}\right),$$

to get

$$\begin{aligned}P_tf(x) &= f\left(\Psi(x,t)\right)\exp\left(-\int_0^t(c+d\Psi(x,s))ds\right)\\ &\quad+\int_x^{\Psi(x,t)}P_{t-\varphi(x,v)}f(v+1)\exp\left(-\int_0^{\varphi(x,v)}(c+d\Psi(x,s))ds\right)\frac{c+dv}{a-bv}dv.\end{aligned}$$

Note that $\Psi(x,t)\le\Psi(a/b-\varepsilon,t)<a/b$, so that $a-bv\neq0$. Since $s\mapsto P_sf(x)$, Ψ, φ and f are smooth, $x\mapsto P_tf(x)\in\mathscr{C}^N([o,a/b-\varepsilon])$. The reasoning holds with the same substitution for $x\ge a/b+\varepsilon$, so that $P_tf\in\mathscr{C}^N(\mathbb{R}_+\backslash\{a/b\})$. Now, if $x>a/b-\varepsilon$, for any $u>0$,

$$\Psi(x,u)+1\ge a/b+1-\varepsilon\ge a/b+\varepsilon,$$

so $x \mapsto P_{t-u}f(\Psi(x,u)+1)$ is smooth. Thus the right-hand side of (4.5.1) is smooth as well and $P_t f \in \mathscr{C}^N(\mathbb{R}_+)$.

Now, let us show that the semigroup generated by $\mathcal{L}$ has bounded derivatives. Note that it is possible to mimic this proof for the example of the WRW treated in Section 4.3.1 when the derivatives of $P_t f$ are not explicit. Let $\mathcal{A}_n f = f^{(n)}$, $\mathcal{J}f(x) = f(x+1) - f(x)$ and $\psi_n(s) = P_{t-s}\mathcal{A}_n P_s f$ for $0 \leq n \leq N$. So, $\psi_n'(s) = P_{t-s}(\mathcal{A}_n\mathcal{L} - \mathcal{L}\mathcal{A}_n)P_s f$. It is clear that $\mathcal{A}_{n+1} = \mathcal{A}_1\mathcal{A}_n$, that $\mathcal{A}_n\mathcal{J} = \mathcal{J}\mathcal{A}_n$ and that

$$\mathcal{L}g(x) = (a-bx)\mathcal{A}_1 g(x) + (c+dx)\mathcal{J}g(x).$$

It is straightforward by induction that

$$\mathcal{A}_n\mathcal{L}g = \mathcal{L}\mathcal{A}_n g - nb\mathcal{A}_n g + nd\mathcal{J}\mathcal{A}_{n-1}g,$$

so the following inequality holds:

$$(\mathcal{A}_n\mathcal{L} - \mathcal{L}\mathcal{A}_n)\, g \leq -nb\mathcal{A}_n g + 2|d|n\|\mathcal{A}_{n-1}g\|_\infty.$$

Hence,

$$\psi_n'(s) \leq -nb\psi_n(s) + 2|d|n\|\mathcal{A}_{n-1}P_s f\|_\infty.$$

In particular, $\psi_1'(s) \leq -b\psi_1(s) + 2d\|f\|_\infty$, so, by Grönwall's inequality,

$$\psi_1(s) \leq \left(\psi_1(0) - \frac{2|d|}{b}\|f\|_\infty\right)e^{-bs} + \frac{2|d|}{b}\|f\|_\infty \leq \|f'\|_\infty + \frac{2d}{b}\|f\|_\infty.$$

Let us show by induction that

$$\psi_n(s) \leq \sum_{k=0}^{n}\left(\frac{2|d|}{b}\right)^{n-k}\|f^{(k)}\|_\infty. \tag{4.5.2}$$

If (4.5.2) is true for some $n \geq 1$ (we denote by K_n its right-hand side), then for all $t < T$, $\psi_n(t) \leq K_n$ and, since $\mathcal{A}_n P_t(-f) = -\mathcal{A}_n P_t f$, $|\psi_n(t)| \leq K_n$, so $\|\mathcal{A}_n P_s f\|_\infty \leq K_n$. Then, we deduce that $\psi_{n+1}'(s) \leq -(n+1)b\psi_{n+1}(s) + 2(n+1)dK_n$. Use Grönwall's inequality once more to have $\psi_{n+1}(s) \leq K_{n+1}$ and achieve the proof by induction. In particular, taking $s = t$ in (4.5.2) provides $\mathcal{A}_n P_t f \leq K_n$ and, since $\mathcal{A}_n P_t(-f) = -\mathcal{A}_n P_t f$, $\mathcal{A}_n P_t f \leq K_n$. As a conclusion, for $n \in \{0, \dots, N\}$,

$$\|(P_t f)^{(n)}\|_\infty \leq \sum_{k=0}^{n}\left(\frac{2|d|}{b}\right)^{n-k}\|f^{(k)}\|_\infty,$$

which concludes the proof when $b > 0$.

The case $b = 0$ is dealt with in a similar way. We use the substitution $\varphi(x,v) = (v-x)/a$ in (4.5.1), which is enough to prove smoothness (this time, $\Psi(x,\cdot)$ is a diffeomorphism for any $x \geq 0$), and it is easy to mimic the proof to obtain the following estimates, for $s \leq t$,

$$|\psi_n(s)| \leq \sum_{k=0}^{n}\frac{n!}{k!}(2|d|T)^{n-k}\|f^{(k)}\|_\infty.$$

□

Proof of Lemma 4.3.8: First, we shall prove that Assumption 4.2.2 holds; let

$$y \in \mathrm{Supp}(\mathscr{L}(y_n^{(l,\delta)})) = [0, \delta\sqrt{n}].$$

Note that $\widetilde{I}_n^0(y), I_n^0(y) \leq 1$ and $\widetilde{I}_n^1(y), I_n^1(y) \leq 0$, so if $y_n^{(l,\delta)} \leq \delta\gamma_{n+1}^{-1} - 1$, then $y_{n+1}^{(l,\delta)} \leq \delta\gamma_{n+1}^{-1}$. For $f \in \mathscr{F}$,

$$\begin{aligned}
|\mathcal{L}_n^{(l,\delta)} f(y) - \mathcal{L}_n f(y)| &\leq \gamma_{n+1}^{-1} \mathbb{E}\left[f(y_{n+1}^{(l,\delta)}) - f(y_{n+1}) \middle| y_n = y_n^{(l,\delta)} = y \right] \\
&\leq \frac{\mathbb{1}_{y \geq \delta\gamma_{n+1}^{-1} - 1}}{\gamma_{n+1}} \Big(p_0(1-\gamma_n y) \left| f(\delta\gamma_{n+1}^{-1}) - f(y + I_n^0(y)) \right| \\
&\quad + \widetilde{p}_0(1-\gamma_n y) \left| f(\delta\gamma_{n+1}^{-1}) - f(y + \widetilde{I}_n^0(y)) \right| \Big) \\
&\leq \frac{\|f'\|_\infty \mathbb{1}_{y \geq \delta\gamma_{n+1}^{-1} - 1}}{\gamma_{n+1}} \left(p_0(1-\gamma_n y) + \widetilde{p}_0(1-\gamma_n y)\right) \leq \frac{y+1}{\delta} \|f'\|_\infty \mathbb{1}_{y \geq \delta\gamma_{n+1}^{-1} - 1} \\
&\leq \frac{(y+1)^2}{\delta^2} \|f'\|_\infty \gamma_{n+1}.
\end{aligned}$$

Using this inequality with (4.3.13), we can explicit the convergence of $\mathcal{L}_n^{(l,\delta)}$ toward $\mathcal{L}$ defined in (4.3.6):

$$\begin{aligned}
|\mathcal{L}_n^{(l,\delta)} f(y) - \mathcal{L} f(y)| &\leq |\mathcal{L}_n^{(l,\delta)} f(y) - \mathcal{L}_n f(y)| + |\mathcal{L}_n f(y) - \mathcal{L} f(y)| \\
&= \chi_3(y)(\|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty)\mathcal{O}(\gamma_n).
\end{aligned} \tag{4.5.3}$$

Note that the notation $\mathcal{O}$ depends here on l and δ, but is uniform over y and f.

Assumption 4.2.3 holds, since it takes into account only the limit process generated by $\mathcal{L}$, and it is a consequence of Proposition 4.3.6: for $n \leq 3$,

$$\|(P_t f)^{(n)}\|_\infty \leq \sum_{k=0}^{n} \left(\frac{2|p_0'(1)|}{p_1(1)} \right)^{n-k} \|f^{(k)}\|_\infty.$$

Now, we shall check a Lyapunov criterion for the chain $(y_n^{(l,\delta)})_{n\geq 0}$, in order to ensure Assumption 4.2.4. Taking $V(y) = \mathrm{e}^{\theta y}$, where (small) $\theta > 0$ will be chosen afterwards, we have, for $n \geq l$ and $y \leq \delta\gamma_n^{-1}$,

$$\begin{aligned}
\mathcal{L}_n^{(l,\delta)} V(y) &\leq \gamma_{n+1}^{-1} \mathbb{E}\left[V((y + I_n(y)) \wedge \delta\sqrt{n}) - V(y) \right] \leq \gamma_{n+1}^{-1} \mathbb{E}\left[V(y + I_n(y)) - V(y) \right] \\
&\leq V(y)\sqrt{n+1}\left(\mathbb{E}[\mathrm{e}^{\theta I_n(y)}] - 1 \right).
\end{aligned}$$

Let $\varepsilon > 0$; we are going to decompose $I_n(y)$. The first term is

$$\begin{aligned}
&\sqrt{n+1}\left(\exp\left(\frac{\sqrt{n+1} - \sqrt{n} - 1}{\sqrt{n}} \theta y \right) - 1 \right) p_1(1-\gamma_n y) \\
&\leq \sqrt{n+1}\left(\frac{\sqrt{n+1} - \sqrt{n} - 1}{\sqrt{n}} \theta y + \frac{1}{2}\left(\frac{\sqrt{n+1} - \sqrt{n} - 1}{\sqrt{n}} \theta y \right)^2 \right) p_1(1-\gamma_n y) \\
&\leq \left(-\alpha_n \theta y + \frac{\alpha_n^2}{2\sqrt{n+1}} \theta^2 y^2 \right) p_1(1-\gamma_n y) \leq \theta y \left(-\alpha_n + \frac{\alpha_n^2}{2}\theta\delta \right) p_1(1-\gamma_n y) \\
&\leq \left(\varepsilon + \left(-1 + \frac{\theta\delta}{2} \right) \right) \theta y \quad \text{for } n \text{ large.}
\end{aligned}$$

where $\alpha_n = \left(1 - \sqrt{n+1} + \sqrt{n}\right)\gamma_n\gamma_{n+1}^{-1}$. There exists $\xi^{(\delta)}$, such that $1-\delta \leq \xi^{(\delta)} \leq 1$ and the second term writes:

$$\sqrt{n+1}\left(\exp\left(\theta + \frac{\sqrt{n+1}-\sqrt{n}-1}{\sqrt{n}}\theta y\right) - 1\right)p_0(1-\gamma_n y) \leq \sqrt{n+1}p_0(1-\gamma_n y)(\mathrm{e}^\theta - 1)$$
$$\leq -\sqrt{n+1}\gamma_n y p_0'(\xi^{(\delta)})(\mathrm{e}^\theta - 1) \leq \left(\varepsilon - (\mathrm{e}^\theta - 1)p_0'(1)\right) y \quad \text{for } n \text{ large.}$$

The third term is negative, and the fourth term writes:

$$\sqrt{n+1}\left(\exp\left(\frac{\theta}{\sqrt{n+1}} + \frac{n-\sqrt{n(n+1)}}{\sqrt{n(n+1)}}\theta y\right) - 1\right)\widetilde{p}_0(1-\gamma_n y)$$
$$\leq \sqrt{n+1}\left(\exp\left(\frac{\theta}{\sqrt{n+1}}\right) - 1\right) \leq \theta + \varepsilon \quad \text{for } n \text{ large.}$$

Hence, there exists some (deterministic) $n_0 \geq l$ such that, for $n \geq n_0$,

$$\mathcal{L}_n^{(l,\delta)}V(y) \leq V(y)\left[\theta + \varepsilon - y\left(p_0'(1)(\mathrm{e}^\theta - 1) - \left(\theta + \frac{\theta\delta}{2}\right)p_1(1) + \epsilon(1+\theta)\right)\right].$$

Then, for $\varepsilon, \delta, \theta$ small enough, there exists $\widetilde{\alpha} > 0$ such that, for $n \geq n_0$ and for any $M \geq \widetilde{(\theta+\epsilon)}\alpha^{-1}$,

$$\mathcal{L}_n^{(l,\delta)}V(y) \leq V(y)(\theta + \varepsilon - \widetilde{\alpha}y) \leq -(\widetilde{\alpha}M - \theta - \varepsilon)V(y) + \widetilde{\alpha}MV(M).$$

Then, Assumption 4.2.4.iii holds with

$$\alpha = \left(p_0'(1)(\mathrm{e}^\theta - 1) - \left(\theta + \frac{\theta\delta}{2}\right)p_1(1) + \epsilon(1+\theta)\right)M - \theta - \varepsilon, \quad \beta = \widetilde{\alpha}MV(M).$$

Finally, checking Assumption 4.2.12 is easy (using (4.5.3) for instance) with $d_2 = 3$, which forces us to set $d = 6$ (since $\Gamma_n\chi_3 \leq M_6\chi_6$). The chain $(y_n^{(l,\delta)})_{n\geq 0}$ satisfying a Lyapunov criterion with $V(y) = \mathrm{e}^{\theta y}$, its moments of order 6 are also uniformly bounded. □

4.5.3 Appendix for the decreasing step Euler scheme

Proof of Lemma 4.3.12: Applying Itô's formula with $x \mapsto |x|^p$, we get

$$\begin{aligned}|\partial_x X_t^x|^p &= 1 + \int_0^t p\left(b'(X_s^x)|\partial_x X_s^x|^p + \frac{p-1}{2}(\sigma'(X_s^x))^2|\partial_x X_s^x|^p\right)ds \\ &\quad + \int_0^t p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s \\ &\leq 1 + C\int_0^t |\partial_x X_s^x|^p ds + \int_0^t p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s, \end{aligned} \tag{4.5.4}$$

where $C = p\|b'\|_\infty + \frac{p(p-1)}{2}\|\sigma'\|_\infty^2$. Let us show that $\int_0^t p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s$ is a martingale. To that end, since $|\partial_x X_t^x|^p$ is non-negative and $(x+y+z)^2 \leq 2(x^2+y^2+z^2)$, we use the Burkholder–Davis–Gundy's inequality so there exists a constant C' such that,

$$
\begin{aligned}
|\partial_x X_t^x|^p &\leq 1 + C\int_0^t \sup_{u\in[0,s]} |\partial_x X_u^x|^p ds + \int_0^t p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s \\
\sup_{u\in[0,t]} |\partial_x X_u^x|^p &\leq 1 + C\int_0^t \sup_{u\in[0,s]} |\partial_x X_u^x|^p ds + \sup_{u\in[0,t]}\int_0^u p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s \\
\mathbb{E}\left[\sup_{u\in[0,t]} |\partial_x X_u^x|^{2p}\right] &\leq 2 + 2C^2 T\int_0^t \mathbb{E}\left[\sup_{u\in[0,s]} |\partial_x X_u^x|^{2p}\right] ds \\
&\quad + 2\mathbb{E}\left[\left(\sup_{u\in[0,t]}\int_0^u p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s\right)^2\right] \\
&\leq 2 + 2C^2 T\int_0^t \mathbb{E}\left[\sup_{u\in[0,s]} |\partial_x X_u^x|^{2p}\right] ds + 2C'\int_0^t \mathbb{E}[\sigma'(X_s^x)^2|\partial_x X_s^x|^{2p}]ds \\
&\leq 2 + 2C^2 T\int_0^t \mathbb{E}\left[\sup_{u\in[0,s]} |\partial_x X_u^x|^{2p}\right] ds \\
&\quad + 2C'\|\sigma'\|_\infty^2\int_0^t \mathbb{E}\left[\sup_{u\in[0,s]} |\partial_x X_u^x|^{2p}\right] ds \\
&\leq 2\exp\left((C^2 T + C'\|\sigma'\|_\infty^2)T\right) \quad \text{by Grönwall's Lemma.}
\end{aligned}
$$

Hence, $\int_0^t p\sigma'(X_s^x)|\partial_x X_s^x|^p dW_s$ is a martingale and, taking the expected values in (4.5.4) and applying Grönwall's lemma once again, we have

$$
\mathbb{E}[|\partial_x X_t^x|^p] \leq \exp\left(\left(p\|b'\|_\infty + \frac{p(p-1)}{2}\|\sigma'\|_\infty^2\right)T\right).
$$

Using Hölder's inequality for $p=2$ completes the case of the first derivative.

Since the following computations are more and more tedious, we choose to treat only the case of the second derivative. Note that $\partial_x^2 X_t^x$ exists and satisfies the following SDE:

$$
\begin{aligned}
\partial_x^2 X_t^x = &\int_0^t \left(b'(X_s^x)\partial_x^2 X_s^x + b''(X_s^x)(\partial_x X_s^x)^2\right) ds \\
&+ \int_0^t \left(\sigma'(X_s^x)\partial_x^2 X_s^x + \sigma''(X_s^x)(\partial_x X_s^x)^2\right) dW_s.
\end{aligned}
$$

Itô's formula provides us the following inequation:

$$
\begin{aligned}
|\partial_x^2 X_t^x|^p \leq\ &C_1\int_0^t |\partial_x^2 X_s^x|^p ds + C_2\int_0^t |\partial_x^2 X_s^x|^{p-1}|\partial_x X_s^x|^2 ds + C_3\int_0^t |\partial_x^2 X_s^x|^{p-2}|\partial_x X_s^x|^4 ds \\
&+ \int_0^t p\Big(|\partial_x^2 X_s^x|^p\sigma'(X_s^x) + |\partial_x^2 X_s^x|^{p-1}\mathrm{sgn}(\partial_x^2 X_s^x)\sigma''(X_s^x)|\partial_x X_s^x|^2\Big) dW_s,
\end{aligned}
$$

with constants C_i depending on $p, \|b'\|_\infty, \|b''\|_\infty, \|\sigma'\|_\infty, \|\sigma''\|_\infty$. The last term proves to be a martingale, with similar arguments as above. We take the expected values, and apply Hölder's inequality twice to find, for $p > 2$,

$$\begin{aligned}
\mathbb{E}\left[|\partial_x^2 X_t^x|^p\right] &\leq C_1 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^p\right] ds + C_2 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^{p-1} |\partial_x X_s^x|^2\right] ds \\
&\quad + C_3 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^{p-2} |\partial_x X_s^x|^4\right] ds \\
&\leq C_1 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^p\right] ds + C_2 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^p\right]^{\frac{p-1}{p}} \mathbb{E}\left[|\partial_x X_s^x|^{2p}\right]^{\frac{1}{p}} ds \\
&\quad + C_3 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^p\right]^{\frac{p-2}{p}} \mathbb{E}\left[|\partial_x X_s^x|^{2p}\right]^{\frac{2}{p}} ds \\
&\leq C_3 e^{C_4 T} + C_1 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^p\right] ds + (C_2 + C_3) e^{C_4 T} \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^p\right]^{\frac{p-1}{p}} ds,
\end{aligned}$$

with $C_4 = 4\|b'\|_\infty + 2(p-1)\|\sigma'\|_\infty^2$. The case $p = 2$ is deduced straightforwardly:

$$\mathbb{E}\left[|\partial_x^2 X_t^x|^2\right] \leq C_3 e^{C_4 T} + C_1 \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^2\right] ds + C_3 e^{C_4 T} \int_0^t \mathbb{E}\left[|\partial_x^2 X_s^x|^2\right]^{\frac{1}{2}} ds.$$

Regardless, since the unique solution of $u = Au + Bu^\alpha$ is

$$u(t) = \left(\left(u(0)^{1-\alpha} + \frac{B}{A}\right) \exp(A(1-\alpha)t) - \frac{B}{A}\right)^{\frac{1}{1-\alpha}},$$

for $A, B > 0, \alpha \in (0,1), u(0) > 0$, we have

$$\begin{aligned}
\mathbb{E}\left[|\partial_x^2 X_t^x|^2\right] &\leq \left(\left(C_2^{\frac{1}{p}} e^{\frac{C_4}{p} T} + \frac{C_2 + C_3}{C_1} e^{C_4 T}\right) e^{\frac{C_1}{p} T} - \frac{C_2 + C_3}{C_1} e^{C_4 T}\right)^p \\
&\leq \left(C_2^{\frac{1}{p}} e^{\frac{C_4}{p} T} + \frac{C_2 + C_3}{C_1} e^{C_4 T}\right)^p e^{C_1 T}.
\end{aligned}$$

The same reasoning for the third derivative achieves the proof. □

Remark 4.5.1 (***Regularity of general diffusion processes***)**:** The quality of approximation of a diffusion process is not completely unrelated to its regularity, see for instance [HHJ15, Theorem 1.3]. In higher dimension, smoothness is generally checked under Hörmander conditions (see e.g. [Hai11, HHJ15]). ♢

BIBLIOGRAPHIE

[ABC$^+$00] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux. 19, 20, 74, 82, 85, 94

[ABG$^+$14] R. Azaïs, J.-B. Bardet, A. Génadot, N. Krell, and P.-A. Zitt. Piecewise deterministic Markov process—recent results. In *Journées MAS 2012*, volume 44 of *ESAIM Proc.*, pages 276–290. EDP Sci., Les Ulis, 2014. 8, 56

[ADF01] C. Andrieu, A. Doucet, and W. J. Fitzgerald. An introduction to Monte Carlo methods for Bayesian data analysis. In *Nonlinear dynamics and statistics (Cambridge, 1998)*, pages 169–217. Birkhäuser Boston, Boston, MA, 2001. 6

[ADGP14] R. Azaïs, F. Dufour, and A. Gégout-Petit. Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes. *Scand. J. Stat.*, 41(4) :950–969, 2014. 8, 23

[AG15] R. Azaïs and A. Genadot. Semi-parametric inference for the absorption features of a growth-fragmentation model. *TEST*, 24(2) :341–360, 2015. 9

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008. 13

[All11] L. J. S. Allen. *An introduction to stochastic processes with applications to biology.* CRC Press, Boca Raton, FL, second edition, 2011. 9

[AM15] R. Azaïs and A. Muller-Gueudin. Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes. *ArXiv e-prints*, June 2015. 23

[AMTU01] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2) :43–100, 2001. 22

[Asm03] S. Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability. 8, 28, 30, 33, 36, 37, 39, 62, 64

[Bak94] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Math.*, pages 1–114. Springer, Berlin, 1994. 19, 20, 74, 82, 85, 94

[BBC16] M. Benaïm, F. Bouguet, and B. Cloez. Ergodicity of inhomogeneous Markov chains through asymptotic pseudotrajectories. *ArXiv e-prints*, January 2016. 73

[BCF15] J.-B. Bardet, A. Christen, and J. Fontbona. Quantitative exponential bounds for the renewal theorem with spread-out distributions. *ArXiv e-prints*, April 2015. 8, 18

[BCG08] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3) :727–759, 2008. 19, 20

[BCG13a] D. Balagué, J. A. Cañizo, and P. Gabriel. Fine asymptotics of profiles and relaxation to equilibrium for growth-fragmentation equations with variable drift rates. *Kinet. Relat. Models*, 6(2) :219–243, 2013. 59

[BCG+13b] J.-B. Bardet, A. Christen, A. Guillin, F. Malrieu, and P.-A. Zitt. Total variation estimates for the TCP process. *Electron. J. Probab.*, 18(10) :1–21, 2013. 9, 18, 28, 53

[BCT08] P. Bertail, S. Clémençon, and J. Tressou. A storage model with random release rate for modeling exposure to food contaminants. *Math. Biosci. Eng.*, 5(1) :35–60, 2008. 5, 15, 25, 26

[BCT10] P. Bertail, S. Clémençon, and J. Tressou. Statistical analysis of a dynamic model for dietary contaminant exposure. *J. Biol. Dyn.*, 4(2) :212–234, 2010. 8, 23

[BD12] H. Biermé and A. Desolneux. A Fourier approach for the level crossings of shot noise processes with jumps. *J. Appl. Probab.*, 49(1) :100–113, 2012. 23, 62

[Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999. 22, 74, 76, 78, 80, 101, 102

[BH96] M. Benaïm and M. W. Hirsch. Asymptotic pseudotrajectories and chain recurrent flows, with applications. *J. Dynam. Differential Equations*, 8(1) :141–176, 1996. 22, 74, 76

[Bil99] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication. 21, 82, 102

[BKKP05] O. Boxma, H. Kaspi, O. Kella, and D. Perry. On/off storage systems with state-dependent input, output, and switching rates. *Probab. Engrg. Inform. Sci.*, 19(1) :1–14, 2005. 9, 65, 66

[BL14] M. Benaïm and C. Lobry. Lotka Volterra in fluctuating environment or ”how switching between beneficial environments can make survival harder”. *ArXiv e-prints*, December 2014. 10, 63

[BLBMZ12] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.*, 17 :no. 56, 14, 2012. 45

[BLBMZ14] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. On the stability of planar randomly switched systems. *Ann. Appl. Probab.*, 24(1) :292–311, 2014. 10, 63

[BLBMZ15] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Qualitative properties of certain piecewise deterministic Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(3) :1040–1075, 2015. 68

[BLPR07] A Bobrowski, T. Lipniacki, K. Pichór, and R. Rudnicki. Asymptotic behavior of distributions of mRNA and protein levels in a model of stochastic gene expression. *J. Math. Anal. Appl.*, 333(2) :753–769, 2007. 10

[BMP+15] F. Bouguet, F. Malrieu, F. Panloup, C. Poquet, and J. Reygner. Long time behavior of Markov processes and beyond. *ESAIM Proc. Surv.*, 51 :193–211, 2015. 49, 90, 93

[Bon95] J.-L. Bon. *Fiabilité des systèmes : Méthodes mathématiques*. Masson, 1995. 6, 26, 63

[Bou15] F. Bouguet. Quantitative speeds of convergence for exposure to food contaminants. *ESAIM Probab. Stat.*, 19 :482–501, 2015. 25, 50, 53, 74

[BR15a] V. Bally and V. Rabiet. Asymptotic behavior for multi-scale PDMP's. Preprint HAL, April 2015. 89

[BR15b] J. Bierkens and G. Roberts. A piecewise deterministic scaling limit of Lifted Metropolis-Hastings in the Curie-Weiss model. *ArXiv e-prints*, September 2015. 10

[Bre10] J.-C. Breton. Regularity of the laws of shot noise series and of related processes. *J. Theoret. Probab.*, 23(1) :21–38, 2010. 62

[CD08] O. L. V. Costa and F. Dufour. Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, 47(2) :1053–1077, 2008. 8

[CDG12] V. Calvez, M. Doumic, and P. Gabriel. Self-similarity in a general aggregation-fragmentation problem. Application to fitness analysis. *J. Math. Pures Appl. (9)*, 98(1) :1–27, 2012. 9, 56, 57, 58

[CGLP12] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012. 81

[CGZ13] P. Cattiaux, A. Guillin, and P.-A. Zitt. Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1) :95–118, 2013. 20

[CH15] B. Cloez and M. Hairer. Exponential ergodicity for Markov processes with random switching. *Bernoulli*, 21(1) :505–536, 2015. 68, 74

[Che04] M.-F. Chen. *From Markov chains to non-equilibrium particle systems.* World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004. 85, 105

[Clo12] B. Cloez. Wasserstein decay of one dimensional jump-diffusions. *ArXiv e-prints*, February 2012. 74

[Clo13] B. Cloez. *Asymptotic behavior of jump processes and applications for branching models.* Theses, Université Paris-Est, June 2013. 17

[CMP10] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8) :1518–1534, 2010. 9, 18, 28, 56

[Cos90] O. L. V. Costa. Stationary distributions for piecewise-deterministic Markov processes. *J. Appl. Probab.*, 27(1) :60–73, 1990. 8

[CT09] S. Clémençon and J. Tressou. Exposition aux risques alimentaires et processus stochastiques : le cas des contaminants chimiques. *J. SFdS*, 150(1) :3–29, 2009. 23

[dA82] A. de Acosta. Invariance principles in probability for triangular arrays of B-valued random vectors and some applications. *Ann. Probab.*, 10(2) :346–373, 1982. 12

[Dav93] M.H.A. Davis. *Markov Models & Optimization*, volume 49. CRC Press, 1993. 1, 7, 50, 57, 88

[DFG09] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3) :897–923, 2009. 15, 80

[DHKR15] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3) :1760–1799, 2015. 9, 23, 56

[DHRBR12] M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM J. Numer. Anal.*, 50(2) :925–950, 2012. 23

[DJG10] M. Doumic Jauffret and P. Gabriel. Eigenelements of a general aggregation-fragmentation model. *Math. Models Methods Appl. Sci.*, 20(5) :757–783, 2010. 56

[DMR04] R. Douc, E. Moulines, and J. S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab.*, 14(4) :1643–1665, 2004. 15

[DMT95] D. Down, S. P. Meyn, and R. L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, 23(4) :1671–1691, 1995. 14, 15

[Duf96] M. Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996. 74

[Ebe11] A. Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20) :1101–1104, 2011. 74

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence. 4, 14, 75, 76

[Fel71] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971. 30

[FGM12] J. Fontbona, H. Guérin, and F. Malrieu. Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.*, 44(4) :977–994, 2012. 10, 18, 63

[FGM15] J. Fontbona, H. Guérin, and F. Malrieu. Long time behavior of telegraph processes under convex potentials. *ArXiv e-prints*, July 2015. 18

[FGR09] A. Faggionato, D. Gabrielli, and M. Ribezzi Crivellari. Stationarity, time–reversal and fluctuation theory for a class of piecewise deterministic Markov processes. *ArXiv e-prints*, February 2009. 64

[For15] G. Fort. Central limit theorems for stochastic approximation with controlled Markov chain dynamics. *ESAIM Probab. Stat.*, 19 :60–80, 2015. 50, 74

[Fou02] N. Fournier. Jumping SDEs : absolute continuity using monotonicity. *Stochastic Process. Appl.*, 98(2) :317–330, 2002. 8

[Gen03] I. Gentil. The general optimal L^p-Euclidean logarithmic Sobolev inequality by Hamilton-Jacobi equations. *J. Funct. Anal.*, 202(2) :591–599, 2003. 22

[GP82] M. Gibaldi and D. Perrier. *Pharmacokinetics, Second Edition*. Drugs and the Pharmaceutical Sciences. Taylor & Francis, 1982. 9, 23

[GPS15] S. Gadat, F. Panloup, and S. Saadane. Regret bounds for Narendra-Shapiro bandit algorithms. *ArXiv e-prints*, February 2015. 49, 91, 93

[GRS96] W. Gilks, S. Richardson, and D. Spiegelhalter. Markov Chain Monte Carlo in Practice, ser. interdisciplinary statistics series, 1996. 6

[Hai10] M. Hairer. Convergence of Markov processes. `http://www.hairer.org/notes/Convergence.pdf`, 2010. 28, 80

[Hai11] M. Hairer. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.*, 135(6-7) :650–666, 2011. 112

[HHJ15] M. Hairer, M. Hutzenthaler, and A. Jentzen. Loss of regularity for Kolmogorov equations. *Ann. Probab.*, 43(2) :468–527, 2015. 112

[HM11] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011. 15, 74, 80

[HMS11] M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2) :223–259, 2011. 74

[HT89] T. Hsing and J. L. Teugels. Extremal properties of shot noise processes. *Adv. in Appl. Probab.*, 21(3) :513–525, 1989. 62

[IJ03] A. M. Iksanov and Z. J. Jurek. Shot noise distributions and selfdecomposability. *Stochastic Anal. Appl.*, 21(3) :593–609, 2003. 62

[Iks13] A. Iksanov. Functional limit theorems for renewal shot noise processes with increasing response functions. *Stochastic Process. Appl.*, 123(6) :1987–2010, 2013. 62

[IMM14] A. Iksanov, A. Marynych, and M. Meiners. Limit theorems for renewal shot noise processes with eventually decreasing response functions. *Stochastic Process. Appl.*, 124(6) :2132–2170, 2014. 62

[IW89] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam ; Kodansha, Ltd., Tokyo, second edition, 1989. 8

[Jac06] M. Jacobsen. *Point process theory and applications*. Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 2006. Marked point and piecewise deterministic processes. 27

[JM86] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. in Appl. Probab.*, 18(1) :20–65, 1986. 21, 82, 104

[Jou07] A. Joulin. Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli*, 13(3) :782–798, 2007. 17

[JS03] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003. 21, 82

[Kal02] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. 4, 74

[KM94] H. Kaspi and A. Mandelbaum. On Harris recurrence in continuous time. *Math. Oper. Res.*, 19(1) :211–222, 1994. 61

[KP11] R. M. Kovacevic and G. C. Pflug. Does insurance help to escape the poverty trap ?—a ruin theoretic approach. *Journal of Risk and Insurance*, 78(4) :1003–1028, 2011. 9

[Kun84] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. In *École d'été de probabilités de Saint-Flour, XII—1982*, volume 1097 of *Lecture Notes in Math.*, pages 143–303. Springer, Berlin, 1984. 95

[KY03] H. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35. Springer, 2003. 50, 74, 83, 87, 95

[Lem05] V. Lemaire. *Estimation récursive de la mesure invariante d'un processus de diffusion.* Theses, Université de Marne la Vallée, December 2005. 74, 94, 95

[Lin86] T. Lindvall. On coupling of renewal processes with use of failure rates. *Stochastic Process. Appl.*, 22(1) :1–15, 1986. 26, 28

[Lin92] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication. 8, 12, 16, 28, 54, 74

[LP02] D. Lamberton and G. Pagès. Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 8(3) :367–405, 2002. 74, 94

[LP08a] D. Lamberton and G. Pagès. How fast is the bandit ? *Stoch. Anal. Appl.*, 26(3) :603–623, 2008. 50

[LP08b] D. Lamberton and G. Pagès. A penalized bandit algorithm. *Electron. J. Probab.*, 13 :no. 13, 341–373, 2008. 49, 50, 74, 86, 87, 89

[LP13a] S. Laruelle and G. Pagès. Randomized urn models revisited using stochastic approximation. *Ann. Appl. Probab.*, 23(4) :1409–1436, 2013. 74

[LP13b] A. Löpker and Z. Palmowski. On time reversal of piecewise deterministic Markov processes. *Electron. J. Probab.*, 18 :no. 13, 29, 2013. 8, 64, 67, 71

[LPT04] D. Lamberton, G. Pagès, and P. Tarrès. When can the two-armed bandit algorithm be trusted ? *Ann. Appl. Probab.*, 14(3) :1424–1454, 2004. 50

[LvL08] A. H. Löpker and J. S. H. van Leeuwaarden. Transient moments of the TCP window size process. *J. Appl. Probab.*, 45(1) :163–175, 2008. 9, 56, 59

[Mal15] F. Malrieu. Some simple but challenging Markov processes. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4) :857–883, 2015. 9, 70

[MH10] A. R. Mesquita and J. P. Hespanha. Construction of lyapunov functions for piecewise-deterministic markov processes. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 2408–2413. IEEE, 2010. 15

[Mon14a] P. Monmarché. Hypocoercive relaxation to equilibrium for some kinetic models. *Kinet. Relat. Models*, 7(2) :341–360, 2014. 25, 74

[Mon14b] P. Monmarché. *Hypocoercivité : approche alternative et application aux algorithmes stochastiques*. Theses, Université Toulouse 3 Paul Sabatier, December 2014. 19

[Mon15] P. Monmarché. On $\mathcal{H}^1$ and entropic convergence for contractive PDMP. *ArXiv e-prints*, April 2015. 20, 25

[MT93a] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993. 14, 19, 28

[MT93b] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3) :518–548, 1993. 14, 61, 74, 80, 96, 97

[MZ16] F. Malrieu and P.-A. Zitt. On the persistence regime for Lotka-Volterra in randomly fluctuating environments. *ArXiv e-prints*, January 2016. 10

[OB83] E. Orsingher and F. Battaglia. Probability distributions and level crossings of shot noise models. *Stochastics*, 8(1) :45–61, 1982/83. 23, 62

[Per07] B. Perthame. *Transport equations in biology.* Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007. 10, 56

[Res92] S. Resnick. *Adventures in stochastic processes.* Birkhäuser Boston Inc., Boston, MA, 1992. 28

[RHK+14] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert, and M. Doumic. Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism. *BMC biology*, 12(1) :1, 2014. 23

[Ric77] J. Rice. On generalized shot noise. *Advances in Appl. Probability*, 9(3) :553–565, 1977. 62

[RKSF13] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov, and F. J. Fabozzi. *The methods of distances in the theory of probability and statistics.* Springer, New York, 2013. 105

[RT15] R. Rudnicki and M. Tyran-Kaminska. Piecewise deterministic Markov processes in biological models. *ArXiv e-prints*, December 2015. 9

[Tal84] D. Talay. Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE and applications. In *Filtering and control of random processes (Paris, 1983)*, volume 61 of *Lecture Notes in Control and Inform. Sci.*, pages 294–313. Springer, Berlin, 1984. 93

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4) :483–509 (1991), 1990. 93

[Vil09] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new. 11, 13, 28

[YZLM14] R. Yvinec, C. Zhuge, J. Lei, and M. C. Mackey. Adiabatic reduction of a model of stochastic gene expression with jump Markov process. *J. Math. Biol.*, 68(5) :1051–1070, 2014. 10, 65